Hybrid framework for the simulation of stochastic chemical kinetics

Journal of Computational Physics 326 (2016) 398–419

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Hybrid framework for the simulation of stochastic chemical kinetics

Andrew Duncan [a,*], Radek Erban [b], Konstantinos Zygalakis [c]

[a] Department of Mathematics, Imperial College, South Kensington Campus, London, SW7 2AZ, United Kingdom
[b] Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, United Kingdom
[c] School of Mathematics, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, United Kingdom

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form 9 August 2016
Accepted 24 August 2016
Available online 30 August 2016

Keywords:
Chemical master equation
Chemical Langevin equation
Jump-diffusion process
Hybrid scheme

ABSTRACT

Stochasticity plays a fundamental role in various biochemical processes, such as cell regulatory networks and enzyme cascades. Isothermal, well-mixed systems can be modelled as Markov processes, typically simulated using the Gillespie Stochastic Simulation Algorithm (SSA) [25]. While easy to implement and exact, the computational cost of using the Gillespie SSA to simulate such systems can become prohibitive as the frequency of reaction events increases. This has motivated numerous coarse-grained schemes, where the "fast" reactions are approximated either using Langevin dynamics or deterministically. While such approaches provide a good approximation when all reactants are abundant, the approximation breaks down when one or more species exist only in small concentrations and the fluctuations arising from the discrete nature of the reactions become significant. This is particularly problematic when using such methods to compute statistics of extinction times for chemical species, as well as simulating non-equilibrium systems such as cell-cycle models in which a single species can cycle between abundance and scarcity. In this paper, a hybrid jump-diffusion model for simulating well-mixed stochastic kinetics is derived. It acts as a bridge between the Gillespie SSA and the chemical Langevin equation. For low reactant reactions the underlying behaviour is purely discrete, while purely diffusive when the concentrations of all species are large, with the two different behaviours coexisting in the intermediate region. A bound on the weak error in the classical large volume scaling limit is obtained, and three different numerical discretisations of the jump-diffusion model are described. The benefits of such a formalism are illustrated using computational examples.

1. Introduction

Biochemical systems with small numbers of interacting components have increasingly been studied in the recent years. Examples include the phage λ lysis-lysogeny decision circuit [5], circadian rhythms [57] and cell cycle [39]. It is this small number of interacting components that makes the appropriate mathematical framework for describing these systems a stochastic one. In particular, the kinetics of the different species is accurately described, under appropriate assumptions, by a continuous-time discrete-space Markov chain. The theory of stochastic processes [23,56] allows the association of the

* Corresponding author.
E-mail addresses: a.duncan@imperial.ac.uk (A. Duncan), erban@maths.ox.ac.uk (R. Erban), k.zygalakis@ed.ac.uk (K. Zygalakis).

http://dx.doi.org/10.1016/j.jcp.2016.08.034
0021-9991/

Markov chain with an underlying master equation, which is a set of ordinary differential equations (ODEs), possible of infinite dimensions, that describe, at each point in time, the probability density of all the different possible states of the system. In the context of biochemical systems this equation is known as the *chemical master equation* (CME).

The high dimensionality of the CME makes it intractable to solve in practice. In particular, with the exception of some very simple chemical systems [36] analytic solutions of the CME are not available. One way to deal with this issue is to resort to stochastic simulation of the underlying Markov chain. The stochastic simulation algorithm (SSA) developed by Gillespie [25] exactly simulates trajectories of the CME as the system evolves in time. The main idea behind this algorithm is that, at each time point, one samples a waiting time to the next reaction from an appropriate exponential distribution, while another draw of a random variable is then used to decide which of the possible reactions will actually occur. For suitable classes of chemically reacting systems, one can sometimes use exact algorithms which, although equivalent to the Gillespie SSA are less computationally intensive. Examples include the Gibson–Bruck Next Reaction Method [24] and the Optimised Direct Method [9]. These algorithms can be further accelerated by using parallel computing, for example, on Graphics Processing Units [43,42].

All the methods described above can only go so far in terms of speeding up the simulations, since even with all the possible speed ups running the SSA can be computationally intensive for realistic problems. One approach to alleviate the computational cost is to employ different approximations on the level of the description of the chemical system. For example, in the limit of large molecular populations, the waiting time becomes, on average, very small and under the law of mass action the time evolution of the kinetics is described by a system of ODEs. This system is known as the *reaction rate equation* which describes, approximately, the time evolution of the mean of the evolving Markov chain. An intermediate regime between the SSA and the reaction rate equation is the one where stochasticity is still important, but there exist a sufficient number of molecules to describe the evolving kinetics by a continuous model. This regime is called the chemical Langevin equation (CLE) [56,27], which is an Itô stochastic differential equation (SDE) driven by a multidimensional Wiener process. In this case the corresponding master equation for the CLE is called the chemical Fokker–Planck equation (CFPE) which is a N-dimensional parabolic partial differential equation, where N is the number of the different chemical species present in the system.

The fact that stochasticity is still present in the description of the chemical system, combined with the fact that the underlying CFPE is more amenable to rigorous analysis than the CME, has made the CLE equation a very popular regime used in applications [51,17,41]. However, while there are benefits to working with the CLE/CFPE, this approximation is only valid in the limit of large system volume and provides poor approximations for systems possessing one or more chemical species with low copy numbers. Furthermore, unlike the SSA/CME which ensures that there is always a positive (or zero) number of molecules in the system, the CFPE and CLE can give rise to negative concentrations, so that the chemical species can attain negative copy numbers. As observed in [52], negative copy numbers can occur even for extremely simple chemical reactions such as $\emptyset \to A$. This can have serious mathematical implications, since the CFPE equation might break down completely, due to regions in which the diffusion tensor is no longer positive definite, which makes the underlying problem ill-posed. On the level of the CFPE, one way to deal with such positivity issues is to truncate the domain and artificially impose no flux-boundary conditions along the domain boundary [11,18,53,17,29,10], which will have a negligible effect on the solution when it is concentrated far away from the boundary. When all chemical species exist in sufficiently high concentration, Dirichlet boundary conditions can also be used if one solves the stationary CFPE as an eigenvalue problem [46]. However, as shown in [15], these artificial boundary conditions can result in significant approximation errors when the solution is concentrated near the boundary. Other alternatives have been proposed to overcome the behaviour of the CLE close to the boundary, either by suppressing reaction channels which may cause negativity near the boundary [13], or by extending the domain of the process to allow exploration in the complex domain [52]. In the latter approach the resulting process, called the Complex CLE will have a positive definite diffusion tensor for all time, thus avoiding such breakdowns entirely. However, this method does not accurately capture the CME behaviour near the boundary, and in areas where the CLE is a poor approximation to the CME, the corresponding Complex CLE will suffer equally.

These issues have motivated a number of hybrid schemes which have been obtained by treating only certain chemical species as continuous variables and the others as discrete [34,20,50]. By doing so, such schemes are able to benefit from the computational efficiency of continuum approximations while still taking into account discrete fluctuations when necessary. Typically such schemes involve partitioning the reactions into “fast” and “slow” reactions, with the fast reactions modelled using a continuum approximation (CLE or the reaction rate equation), while using Markov jump process to simulate the discrete reactions. Chemical species which are affected by fast reactions are then modelled as continuous variables while the others are kept discrete. Since the reaction rate depends on the state, it is possible that some fast reactions become slow and vice versa. This is typically accounted for by periodically repartitioning the reactions. Based on this approach, a number of hybrid models have been proposed, such as [31,12], which couple deterministic reaction-rate equations for the fast reactions with Markov jump dynamics for the slow, resulting in a piecewise-deterministic Markov process for the entire system. Error estimates for such systems, in the large volume limit, were carried out in [38]. Similar methods have been proposed, such as [32] and more recently [54]. Other hybrid schemes [33,49] also involve a similar partition into slow and fast species, however the evolution of the slow species is obtained by solving the CME directly, coupled to a number of reaction-rate equations for the fast reactions. The hybrid system is thus reduced to a system of ODEs. An error analysis of these schemes was carried out in [37].

In this paper, we propose a hybrid scheme which uses Langevin dynamics to simulate fast reactions coupled with jump/SSA dynamics to simulate reactions in which the discreteness cannot be discounted. Thus, unlike the previously proposed models, both the continuous and discrete parts of the model are described using a stochastic formulation. Moreover, our scheme does not explicitly keep track of fast and slow reactions, but rather, the process will perform Langevin dynamics in regions of abundance, jump dynamics in regions where one of the involved chemical species is in small concentrations, and a mixture of both in intermediate regions. The resulting process thus becomes a jump-diffusion process with Poisson distributed jumps. The preference of jump over Langevin dynamics is controlled for each individual reaction by means of a *blending* function which is chosen to take value 1 in regions of low concentration, 0 in regions where all involved chemical species are abundant, and smoothly interpolates in between. The choice of each blending region will depend on the reaction rate associated with the given reaction. The region should be generally chosen so that the resulting propensity is large in the continuum region and small in the discrete region. Hybrid models for chemical dynamics involving both jump and diffusive dynamics have been previously studied in various contexts. Recently, a method [22] based on a similar coupling of SSA and Langevin dynamics was proposed. The authors introduce a partition of reactions into fast and slow reactions, applying the diffusion approximation to the fast reactions to obtain a jump-diffusion process. Based on an a-posteriori error estimator the algorithm periodically repartitions the species accordingly. By introducing the blending region our approach no longer requires periodic repartitioning. Other works which have considered hybrid schemes based on jump-diffusion dynamics include [4]. In [19,20] a hybrid scheme based on a similar domain decomposition idea was proposed for simulating spatially-extended stochastic-reaction diffusion models. In one part of the domain a SDE was used to simulate the position of the particles and on the other part a compartment-based jump process for diffusion was used. These two domains were separated by a sharp interface, where corrections to the transition probabilities at the interface were applied to ensure that probability mass was transferred between domains. While such a direct matching between continuum and discrete fluxes at the interface can accurately simulate systems having only reactions with unit jumps, for systems possessing jumps of length 2 or higher, such a direct coupling would cause non-physical results. This scenario is analogous to *ghost forces* which arise in quasi-continuum methods used in the multiscale modelling of materials [6]. Overlap regions are also necessary for coupling Brownian dynamics (SDEs) with mean-field partial differential equations [21].

The paper is organised as follows. In Section 2 after reviewing the CME/SSA and CLE/CFPE formalisms we introduce blending functions and the hybrid jump-diffusion formalism. In Section 3 we derive weak error bounds for the hybrid scheme in the limit of large volume, and in particular show that the hybrid scheme does not perform worse than the CLE in this regime. In Section 4 we describe three possible discretisations of the process, which can be used in practise to simulate the jump-diffusion process. A number of numerical experiments which demonstrate the use of the hybrid scheme are detailed in Sections 5.1, 5.2 and 5.3.

2. Preliminaries

Consider a biochemical network of N chemical species interacting via R reaction channels within an isothermal reactor of fixed volume V. For $i = 1, 2, \dots, N$, denote by $X_i(t)$ the number of molecules of species S_i at time t, and let $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_N(t))$. Under the assumption that the chemical species are well-mixed it can be shown [26] that $\mathbf{X}(t)$ is a continuous time Markov process. When in state $\mathbf{X}(t)$, the j-th reaction gives rise to a transition $\mathbf{X}(t) \to \mathbf{X}(t) + \boldsymbol{\nu}_j$ with exponentially distributed waiting time with inhomogeneous rate $\lambda_j(\mathbf{X}(t))$, where $\lambda_j(\cdot)$ and $\boldsymbol{\nu}_j \in \mathbb{Z}^N$ denote the propensity and stoichiometric vector corresponding to the j-th reaction, respectively. More specifically, each reaction is of the form

$$\mu_{r1}X_1 + \mu_{r2}X_2 + \dots \mu_{rN}X_N \xrightarrow{k_r} \mu_{r1}{}'X_1 + \mu_{r2}{}'X_2 + \dots \mu_{rN}{}'X_N,$$

where $r = 1, 2, \dots, R$, and μ_{ri}, $\mu_{r1}{}' \in \mathbb{N} = \{0, 1, 2, \dots\}$, for $i = 1, 2, \dots, N$. Let us denote $\boldsymbol{\mu}_r = (\mu_{r1}, \mu_{r2}, \dots, \mu_{rN})$ and $\boldsymbol{\mu}'_r = (\mu'_{r1}, \mu'_{r2}, \dots, \mu'_{rN})$. The stoichiometric vectors $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \dots, \boldsymbol{\nu}_R$ are then given by

$$\boldsymbol{\nu}_r = \boldsymbol{\mu}'_r - \boldsymbol{\mu}_r$$

and describe the net change in molecular copy numbers which occurs during the r-th reaction. Under the assumption of mass action kinetics, the propensity function λ_r for the r-th reaction is

$$\lambda_r(\mathbf{x}) = k_r \prod_{j=1}^{N} \frac{x_j!}{(x_j - \mu_{rj})!},$$

where x_j is the number of molecules of the j-th species, while we also assume that $n! = 1$ if $n \leq 0$, to simplify notation. Within the interval $[t, t + \mathrm{d}t)$, we update $\mathbf{X}(t) \to \mathbf{X}(t) + \boldsymbol{\nu}_j$, with probability $\lambda_j(\mathbf{X}(t))\,\mathrm{d}t + o(\mathrm{d}t)$. The process $\mathbf{X}(t)$ can thus be expressed as the sum of R Poisson processes with inhomogeneous rates $\lambda_j(\mathbf{X}(t))$. As noted in [27,44], $\mathbf{X}(t)$ can be expressed as a random time change of unit rate Poisson processes,

$$\mathbf{X}(t) = \mathbf{X}(\mathbf{0}) + \sum_{r=1}^{R} P_r\left(\int_0^t \lambda_r(\mathbf{X}(\mathbf{s}))ds\right)\boldsymbol{\nu}_r, \tag{1}$$

where P_r are independent unit-rate Poisson processes. This is a continuous time Markov process with infinitesimal generator

$$\mathcal{L}_0 f(\mathbf{x}) = \sum_{r=1}^{R} \lambda_r(\mathbf{x})(f(\mathbf{x}+\boldsymbol{\nu}_r) - f(\mathbf{x})). \tag{2}$$

The classical method for sampling realisations of $\mathbf{X}(t)$ is the Gillespie SSA [25]. Given the current state $\mathbf{X}(t)$ at time t, the time of next reaction $t+\tau$ and state $\mathbf{X}(t+\tau)$ are sampled as follows:

1. Let $\lambda_0 = \sum_{r=1}^{R} \lambda_r(\mathbf{X}(t))$.
2. Sample $\tau \sim -\log(u)/\lambda_0$, where $u \sim U[0,1]$.
3. Choose the next reaction r with probability $\lambda_r(\mathbf{X}(t))/\lambda_0$, where $r = 1, 2, \ldots, R$.
4. $\mathbf{X}(t+\tau) = \mathbf{X}(t) + \boldsymbol{\nu}_r$.

We note that in advancing the system from time t to time $t+\tau$ one needs to generate two random numbers each time. Based on the time changed representation (1) one can derive an alternative algorithm, known as the *Next Reaction Method* of Gibson and Bruck [24]. Indeed, for a fixed realisation of each unit rate Poisson process $P_1, P_2, \ldots, P_R$, define $F_r(t)$ to be the last jump time of P_r before time t. Then for each r, the next jump time of P_r after time $F_r(t)$ will be distributed as $F_r(t) - \log u$, where $u \sim U[0,1]$. Clearly, as the process $X(t)$ evolves, the r-th reaction will then occur at $t+\tau_r$ satisfying

$$T_r(t+\tau_r) = F_r(t) - \log u, \qquad \text{where} \qquad T_r(t) = \int_0^t \lambda_r(\mathbf{X}(s))\,\mathrm{d}s.$$

This provides the basis of the Next Reaction method. Suppose we are at time t, the next reaction will occur at $t+\tau_{min}$ for

$$\tau_{min} = \underset{r\in\{1,2,\ldots,R\}}{\operatorname{argmin}} \{\tau_r : F_r(t) - \log u_r = T_r(t+\tau_r)\},$$

where $u_r \sim U[0,1]$ are independently distributed random numbers. Noting that the values of the propensities do not change within $[t, t+\tau_{min})$ we have $T_r(t+\tau) = T_r(t) + \tau_r \lambda_r(\mathbf{X}(t))$, so that the next reaction time is given by

$$\tau_{min} = \underset{r\in\{1,2,\ldots,R\}}{\operatorname{argmin}} \left\{ \frac{F_r(t) - \log u_r - T_r(t)}{\lambda_r(\mathbf{X}(t))} \right\},$$

at which time the reaction that occurs is the one for which $\tau_r = \tau_{min}$. This leads to the following exact algorithm for sampling realisations of $\mathbf{X}(t)$.

1. Set the initial number of molecules of each species, set $t = 0$.
2. Calculate the propensity function λ_r for each reaction.
3. Generate R independent random numbers $u_r \sim U[0,1]$.
4. Set $F_r = -\log(u_r)$ and $T_r = 0$ for each $r = 1, 2, \ldots, R$.
5. Set $\tau_r = (F_r - T_r)/\lambda_r$ for each $r = 1, 2, \ldots, R$.
6. Set $\tau_{min} = \min_r\{\tau_r\}$ and let μ be the reaction for which this minimum is realised.
7. Set $t = t + \tau_{min}$ and update the number of each molecular species according to reaction μ.
8. For each k, set $T_r = T_r + \lambda_r \tau_{min}$, and for the reaction μ, let $u \sim U[0,1]$ and set $F_r = F_r - \log u$.
9. Recalculate the propensity functions λ_r.
10. Return to step 5 or quit.

This algorithm was introduced by Gibson and Bruck [24] who additionally proposed the introduction of an indexed priority queue to efficiently search for the minimum required in step 6, along with a dependency graph structure to efficiently update propensity values in step 9. This makes it less computationally intensive from the Gillespie SSA when simulating systems with many reaction channels [9].

2.1. Diffusion approximation

For $r = 1, 2, \ldots, R$ define $\widetilde{\lambda}_r(\mathbf{x})$ to be a smooth, non-negative extension of $\lambda_r(\mathbf{x})$ from $\mathbb{N}^N$ to $\mathbb{R}^N$ (the precise conditions on this extension are given in Section 3). Given the extended propensities, a commonly used approximation of (1) is the CLE, given by the following Itô SDE

$$\mathrm{d}\mathbf{Y}(t) = \sum_{r=1}^{R} \boldsymbol{\nu}_r \widetilde{\lambda}_r(\mathbf{Y}(t))\,\mathrm{d}t + \sum_{r=1}^{R} \boldsymbol{\nu}_r \sqrt{\widetilde{\lambda}_r(\mathbf{Y}(t))}\,\mathrm{d}W_r(t), \tag{3}$$

where $W_r(t)$ are mutually independent standard Brownian motions. This diffusion approximation is valid in the large volume regime, where all species exist in abundance, and all reactions occur frequently, see for example [27,48]. More precisely, one can show strong convergence of $\mathbf{Y}(t)$ to $\mathbf{X}(t)$ over finite time intervals $[0, T]$, see [2,22]. Clearly, the lifting of $\{\lambda_r\}_{r=1}^R$ to $\{\widetilde{\lambda}_r\}_{r=1}^R$ is not unique, and different extensions will give rise to different diffusion approximations. However, as we shall see in Section 3, in the classical large volume rescaling, the dynamics of the process will be largely determined by the value of the propensities on the rescaled grid $\varepsilon\mathbb{N}^N$, and indeed, subject to the extension satisfying a number of assumptions, different extensions will lead to the diffusion approximation having weak error of the same order.

2.2. The hybrid scheme

In this section, we introduce a jump-diffusion process which provides an approximation which is intermediate between the Gillespie SSA and CLE by introducing a series of *blending* functions $\beta_1, \beta_2, \ldots, \beta_R$ which are used to blend the dynamics linearly between the SSA jump process and the CLE. More specifically, given R smooth functions $\beta_r : \mathbb{R}^d \to [0, 1]$, $r = 1, 2, \ldots, R$ we consider the following Itô jump-diffusion equation

$$\begin{aligned}\mathbf{Z}(t) = \mathbf{Z}(0) &+ \sum_{r=1}^{R} P_r\left(\int_0^t \beta_r(\mathbf{Z}(s))\,\lambda_r(\llbracket \mathbf{Z}(s)\rrbracket)\,\mathrm{d}s\right)\boldsymbol{\nu}_r \\ &+ \sum_{r=1}^{R} \boldsymbol{\nu}_r \int_0^t \Big(1-\beta_r(\mathbf{Z}(s))\Big)\widetilde{\lambda}_r(\mathbf{Z}(s))\,\mathrm{d}s \\ &+ \sum_{r=1}^{R} \boldsymbol{\nu}_r \int_0^t \sqrt{\Big(1-\beta_r(\mathbf{Z}(s))\Big)\widetilde{\lambda}_r(\mathbf{Z}(s))}\,\mathrm{d}W_r(s),\end{aligned} \tag{4}$$

where $\{W_r\}_{r=1}^R$ and $\{P_r\}_{r=1}^R$ are standard Wiener and Poisson processes, respectively, all mutually independent, and $\llbracket \mathbf{x}\rrbracket$ is the closest point in the lattice $\mathbb{Z}^N$ to $\mathbf{x} \in \mathbb{R}^N$. Thus, (4) describes a jump-diffusion Markov process with infinitesimal generator $\mathcal{G}$ defined by

$$\begin{aligned}\mathcal{G}f(\mathbf{z}) = &\sum_{r=1}^{R} \beta_r(\mathbf{z})\lambda_r(\llbracket \mathbf{z}\rrbracket)\,[f(\mathbf{z}+\boldsymbol{\nu}_r) - f(\mathbf{z})] + \sum_{r=1}^{R}(1-\beta_r(\mathbf{z}))\,\widetilde{\lambda}_r(\mathbf{z})\,\boldsymbol{\nu}_r\cdot\nabla f(\mathbf{z}) \\ &+ \frac{1}{2}\sum_{r=1}^{R}(1-\beta_r(\mathbf{z}))\,\widetilde{\lambda}_r(\mathbf{z})\,(\boldsymbol{\nu}_r\otimes\boldsymbol{\nu}_r) : \nabla\nabla f(\mathbf{z}),\end{aligned} \tag{5}$$

for all $f \in C_0^2(\mathbb{R}^N)$, where $\otimes$ stands for the tensor product, and $A : B = \mathrm{trace}(A^T B)$ for square matrices A and B. The first term on the right hand side of (5) captures the jump behaviour of the process, while the remaining two terms encode the effect of the diffusive dynamics.

We see that in regions where $\beta_r(x) = 1$ the dynamics of the r-th reaction is modelled by a pure jump process. Conversely, when $\beta_r(x) = 0$ the dynamics are purely diffusive, corresponding to CLE dynamics. In intermediate regions where $0 < \beta_r(x) < 1$ we obtain a mixture of the two. The rationale is to choose β_r to be 0 in regions where the CLE provides a valid approximation of the biochemical system, and β_r to be 1 in regions where the diffusion approximation breaks down, i.e. in regions where the concentrations of certain species are low and the discrete behaviour becomes significant. An example of a trajectory is shown in Fig. 1, where the blue line depicts diffusive dynamics, while the red lines indicate jumps. We note that the process $\mathbf{Z}(t)$ can still attain negative (and thus non-physical) states, however, $\llbracket \mathbf{Z}(t)\rrbracket$ is always non-negative. A natural interpretation is that one should consider the cell $\llbracket \mathbf{Z}(t)\rrbracket$ as the actual observed dynamics of the system, and the state of the underlying process $\mathbf{Z}(t)$ is a hidden Markov model which is not observed directly.

2.3. Choice of blending functions

The blending functions $\beta_1(x), \beta_2(x), \ldots, \beta_R(x)$ are to be considered as simulation parameters which are chosen to balance the trade-off between the computational cost of using the SSA and the error arising from the diffusion approximation. Generally chosen so that Langevin dynamics are only used in regions where the reactions are considered fast. Since not all species are involved in every reaction, it is natural to choose each blending function differently. For a single species system, a natural choice of blending function is the following piecewise linear function

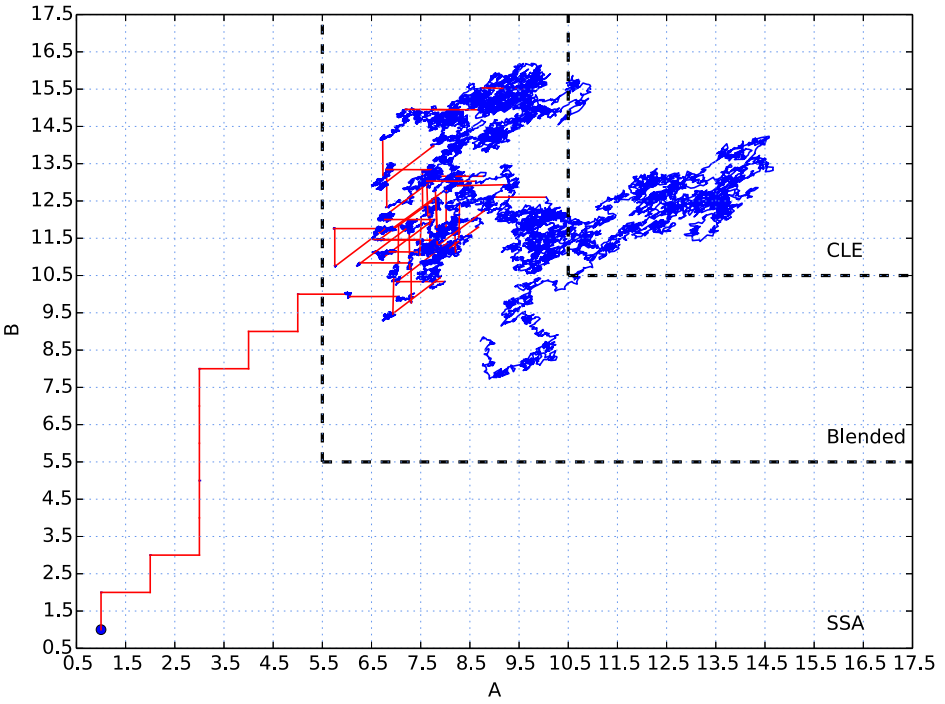


Fig. 1. A typical trajectory for hybrid approximation of the two-species chemical system described in (21) starting from $(A, B) = (1, 1)$. Red lines denote jump dynamics while blue lines CLE dynamics. The black lines demarcate the "blending" region. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\beta(x, I_1, I_2) = \begin{cases} 1, & \text{if } x \leq I_1, \\ \dfrac{I_2 - x}{I_2 - I_1}, & \text{if } I_1 < x < I_2, \\ 0, & \text{if } x \geq I_2, \end{cases}$$

where $0 < I_1 < I_2$ are the boundaries between the different regions. With this choice of $\beta(\cdot, I_1, I_2)$, the hybrid process will perform purely jump dynamics for $0 < \mathbf{Z}(t) < I_1$, purely diffusive dynamics for $I_2 < \mathbf{Z}(t) < \infty$, and jump-diffusion in between.

For chemical systems with N species we can construct blending functions for each reaction as follows. Let S_r be the set of chemical species involved in the r-th reaction (both as reactants, and products of the reaction). Then we can define $\beta_r(\mathbf{x})$, $r = 1, 2, \ldots, R$, as follows

$$\beta_r(\mathbf{x}) = 1 - \prod_{n \in S_r} (1 - \beta(x_n, I_1^n, I_2^n)), \tag{6}$$

where $I_1^n < I_2^n$, $n = 1, 2, \ldots, N$, are the boundaries for each individual chemical species. With this choice of blending function, when $\mathbf{Z}(t)$ is in a state where one of the species involved in the r-th reaction is not abundant, the hybrid process will blend between jumps and diffusion to simulate the corresponding reaction. The choice of the boundaries I_1^n and I_2^n is important to correctly delineate between discrete and continuous behaviour. As will be seen in the numerical experiments in Section 5.2, the accuracy of the scheme is dependent on the width of the blending region. If the blending region $I_2^n - I_1^n$ is small (i.e. close to 1), the step size for the discretised CLE must be accordingly decreased to maintain constant error.

Remark 2.1. While (6) is a natural blending function for the typical biochemical systems where the diffusion approximation breaks down only close to the boundary of the positive orthant, for constrained systems, one must have jump dynamics near all the boundaries of the system. A typical example is the reversible isomerisation model

$$A \underset{k_2}{\overset{k_1}{\rightleftarrows}} B,$$

which can be modelled as a single species birth–death process for A with birth rate $\lambda(m) = k_2(M - m)$ and death rate $\mu(m) = k_1 m$, where M is the total molecule count. For such systems, an appropriate blending function would take value 1 in neighbourhoods of both $m = 0$ and $m = M$.

3. Derivation and consistency of the hybrid scheme

In this section we shall make explicit the regime in which the hybrid scheme correctly captures the dynamics of the original process, and subsequently derive weak error estimates for the expectation of observables at a finite time $T > 0$. Let $\mathbb{L} := \mathbb{N}^N$ be the lattice of possible states. For most biochemical systems, we can make the following natural assumptions.

Assumption 3.1. The propensity functions satisfy $\lambda_j(x) \geq 0$ for all $x \in \mathbb{L}$, and $\lambda_j(x) = 0$ if $x \in \mathbb{L}$ and $x + \boldsymbol{\nu}_j \in \mathbb{Z}^N \setminus \mathbb{L}$.

In particular, we are ensured that the jump process $\mathbf{X}(t)$ never attains a negative state. Chemical Langevin dynamics $\mathbf{Y}(t)$ are only a valid approximation of $\mathbf{X}(t)$ in the large volume limit. To study this regime, we introduce a system size $V \gg 1$ which can be viewed as the (dimensionless) volume of the reactor. Writing $\varepsilon = V^{-1}$, we then rewrite the molecular copy number $\mathbf{X}(t)$ as $\varepsilon^{-1}\mathbf{X}^\varepsilon(t)$ where $\mathbf{X}^\varepsilon(t)$ will be the vector of concentrations of each chemical species. We shall assume that each rate constant k_r satisfies

$$k_r = d_r\, \varepsilon^{-z_r}, \quad \text{where } d_r > 0 \quad \text{and} \quad z_r = -1 + \sum_{k=1}^{K} \mu_{rk}.$$

Given this scaling assumption, we can always write the propensity for the r-th reaction, $r = 1, 2, \ldots, R$, as

$$\lambda_r\left(\frac{\mathbf{x}}{\varepsilon}\right) = \frac{1}{\varepsilon}\,\lambda_r^\varepsilon(\mathbf{x}), \quad \text{for} \quad \mathbf{x} \in \varepsilon\mathbb{L},$$

where $\lambda_r^\varepsilon(\mathbf{x})$ is $O(1)$ with respect to ε. Using (2), the generator of the rescaled process $\mathbf{X}^\varepsilon(t)$ is given as follows:

$$\mathcal{L}^\varepsilon f(\mathbf{x}) = \frac{1}{\varepsilon}\sum_{r=1}^{R} \lambda_r^\varepsilon(\mathbf{x})(f(\mathbf{x} + \varepsilon\boldsymbol{\nu}_r) - f(\mathbf{x})), \quad \text{for} \quad \mathbf{x} \in \varepsilon\,\mathbb{L}. \tag{7}$$

We now introduce the hybrid jump-diffusion scheme. To do so we must extend propensities $\lambda^\varepsilon(\mathbf{x})$ from the discrete lattice $\varepsilon\mathbb{L}$ to $\widetilde{\lambda}^\varepsilon(\mathbf{x})$ on the continuous space $\mathbb{R}^N$. We shall make the following assumptions.

Assumption 3.2. The following properties hold for the extended propensities $\widetilde{\lambda}_j^\varepsilon(x)$:

1. They are non-negative, and lie in $C^3(\mathbb{R}^N)$.
2. They are bounded, uniformly with respect to ε, and the same applies for their mixed derivatives up to order 3.
3. For each j, $\widetilde{\lambda}_j^\varepsilon(x)$ is zero outside a bounded domain Ω^ε.

Remark 3.1. As it stands this assumption will not hold for general chemical reacting systems. Under suitable conditions on the propensities it is possible to replace this assumption with a localisation result showing that the probability of $\mathbf{X}^\varepsilon(t)$ escaping the bounded set within time $[0, T]$ is exponentially small. We shall avoid this approach for simplicity, simply noting that one can always ensure this assumption by setting propensities to zero outside a fixed bounded region.

Remark 3.2. A $C^3(\mathbb{R}^N)$ extension of the propensities satisfying Assumption 3.2 is always possible. Indeed, for each j, set $\widetilde{\lambda}_j$ to be zero in $\mathbb{R}^N \setminus \Omega^\varepsilon$. Then one can extend the value of the propensities to Ω^ε by transfinite interpolation, see [28,7].

Remark 3.3. Such an extension may result in propensities which differ from the "standard" propensities typically used for the CLE. In particular, propensities of the form $k_1 x(x-1)$ must be modified so as to remain non-negative. Such an explicit construction of extended propensities for unimolecular and bimolecular reactions of a single species can be found in [40, Example 4.7-4.8].

Using the extended propensities, one can extend the Markov jump process $\mathbf{X}^\varepsilon(t)$ to take initial conditions $\mathbf{X}^\varepsilon(0) \in \Omega^\varepsilon$. The infinitesimal generator of $\mathbf{X}^\varepsilon(t)$ is the natural extension of (7), also denoted by $\mathcal{L}^\varepsilon$ defined by

$$\mathcal{L}^\varepsilon f(\mathbf{x}) = \frac{1}{\varepsilon}\sum_{r=1}^{R} \widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[f(\mathbf{x} + \varepsilon\boldsymbol{\nu}_r) - f(\mathbf{x})\right], \quad \text{for all } f \in C_0(\mathbb{R}^N).$$

For a fixed observable $g \in C^3(\mathbb{R}^N)$, define the value function $u^\varepsilon : [0, T] \times \mathbb{R}^N \to \mathbb{R}$:

$$u^\varepsilon(t, \mathbf{x}) = \mathbb{E}\left[g(\mathbf{X}^\varepsilon(t)) \,|\, \mathbf{X}^\varepsilon(0) = \mathbf{x}\right].$$

Then $u^\varepsilon(t, \mathbf{x})$ can be expressed as the unique solution of the Backward Kolmogorov equation [23]

$$\begin{aligned} \partial_t u^\varepsilon(t,\mathbf{x}) &= \mathcal{L}^\varepsilon u^\varepsilon(t,\mathbf{x}), \quad \text{for } (t,\mathbf{x}) \in (0,T)\times\mathbb{R}^N, \\ u^\varepsilon(0,\mathbf{x}) &= g(\mathbf{x}), \quad \text{for } \mathbf{x}\in\mathbb{R}^N. \end{aligned} \tag{8}$$

For any fixed $\mathbf{x}\in\mathbb{R}^N$, equation (8) can be viewed as an infinite linear ODE on the translated lattice $\mathbf{x}+\varepsilon\mathbb{L}$. By Assumption 3.2 the propensity is only non-zero for finitely many terms, thus the dynamics is characterised by a finite linear system of ODEs. Existence and uniqueness of a solution $u^\varepsilon(t,\mathbf{x})$ in $C^1[0,T]$ follows immediately. Moreover, if $g(\mathbf{x})$ is locally bounded, then so is $u^\varepsilon(t,\mathbf{x})$. Clearly, for $\mathbf{x}\in\mathbb{R}^N$ such that $\widetilde{\lambda}_r^\varepsilon(\mathbf{x})=0$ for all r we have $u^\varepsilon(t,\mathbf{x})=g(\mathbf{x})$, for all $t\in[0,T]$.

Moreover, using Assumption 3.2, it follows that $u^\varepsilon(t,\cdot)\in C^3(\mathbb{R}^N)$, such that the mixed derivatives can be expressed as the unique solutions of the following equations, where ∂_i, ∂_{ij} and ∂_{ijk} denote first, second and third spatial derivatives with respect to the variables x_i, x_j, x_k, $\forall i,j,k\in\{1,2,\dots,N\}$.

$$\partial_t\partial_i u^\varepsilon(t,\mathbf{x}) - \mathcal{L}^\varepsilon\partial_i u^\varepsilon(t,\mathbf{x}) = \frac{1}{\varepsilon}\sum_{r=1}^R \partial_i\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)-u^\varepsilon(t,\mathbf{x})\right], \tag{9}$$

$$\begin{aligned} \partial_t\partial_{ij} u^\varepsilon(t,\mathbf{x}) - \mathcal{L}^\varepsilon\partial_{ij} u^\varepsilon(t,\mathbf{x}) &= \frac{1}{\varepsilon}\sum_{r=1}^R \partial_{ij}\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)-u^\varepsilon(t,\mathbf{x})\right] \\ &\quad + \frac{1}{\varepsilon}\sum_{a,b\in\Pi_2}\sum_{r=1}^R \partial_a\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[\partial_b u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)-\partial_b u^\varepsilon(t,\mathbf{x})\right], \\ \partial_t\partial_{ijk} u^\varepsilon(t,\mathbf{x}) - \mathcal{L}^\varepsilon\partial_{ijk} u^\varepsilon(t,\mathbf{x}) &= \frac{1}{\varepsilon}\sum_{r=1}^R \partial_{ijk}\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)-u^\varepsilon(t,\mathbf{x})\right] \\ &\quad + \frac{1}{2\varepsilon}\sum_{a,b,c\in\Pi_3}\sum_{r=1}^R \partial_{ab}\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[\partial_c u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)-\partial_c u^\varepsilon(t,\mathbf{x})\right] \\ &\quad + \frac{1}{2\varepsilon}\sum_{a,b,c\in\Pi_3}\sum_{r=1}^R \partial_a\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[\partial_{bc} u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)-\partial_{bc} u^\varepsilon(t,\mathbf{x})\right], \end{aligned}$$

where Π_2 and Π_3 denote the set of permutations of $\{i,j\}$ and $\{i,j,k\}$, respectively.

3.1. The hybrid scheme in the large-volume limit

Having extended the propensities for the discrete process so as to take arbitrary values in $\mathbb{R}^N$, and obtained the equations that the first, second and third order derivatives satisfy, we are now ready to study the weak error that arises from taking the hybrid approximation in the large-volume limit. In Lemma 3.4 we show that spatial derivatives of u^ε are bounded uniformly with respect to ε. This result will be required in Theorem 3.5 to bound the remainder terms in the weak error between the jump and hybrid processes.

Lemma 3.4. *Given $p,q,r=1,2,\dots,R$, and $t\in[0,T]$ define the following scalar quantities*

$$A_r^\varepsilon(t,\mathbf{x}) = \boldsymbol{\nu}_r\cdot\nabla u^\varepsilon(t,\mathbf{x}) = \sum_{i=1}^N \nu_{r,i}\partial_i u^\varepsilon(t,\mathbf{x}),$$

$$B_{p,q}^\varepsilon(t,\mathbf{x}) = (\boldsymbol{\nu}_p\otimes\boldsymbol{\nu}_q):\nabla\nabla u^\varepsilon(t,\mathbf{x}) = \sum_{i,j=1}^N \nu_{p,i}\,\nu_{q,j}\,\partial_{ij}u^\varepsilon(t,\mathbf{x}),$$

$$C_{p,q,r}^\varepsilon(t,\mathbf{x}) = (\boldsymbol{\nu}_p\otimes\boldsymbol{\nu}_q\otimes\boldsymbol{\nu}_r):\nabla\nabla\nabla u^\varepsilon(t,\mathbf{x}) = \sum_{i,j,k=1}^N \nu_{p,i}\,\nu_{q,j}\,\nu_{r,k}\,\partial_{ijk}u^\varepsilon(t,\mathbf{x}).$$

Then, there exists constants K_1, K_2 and K_3 and C_1, C_2, and C_3 independent of ε such that

$$\|A_r^\varepsilon(t,\cdot)\|_\infty \le C_1 e^{K_1 T}, \quad \|B_{p,q}^\varepsilon(t,\cdot)\|_\infty \le C_2 e^{K_2 T}, \quad \|C_{p,q,r}^\varepsilon(t,\cdot)\|_\infty \le C_3 e^{K_3 T}, \tag{10}$$

for $t\in[0,T]$, where the K_1, K_2 and K_3 depend on

$$\left\|\widetilde{\lambda}_r^\varepsilon\right\|_{C^1(\mathbb{R}^N)}, \quad \left\|\widetilde{\lambda}_r^\varepsilon\right\|_{C^2(\mathbb{R}^N)} \quad \text{and} \quad \left\|\widetilde{\lambda}_r^\varepsilon\right\|_{C^3(\mathbb{R}^N)},$$

respectively, which by Assumption 3.2 are bounded independently of ε.

Proof. Using (9), we have

$$\partial_t A_q^\varepsilon(t,\mathbf{x}) - \mathcal{L}^\varepsilon A_q^\varepsilon(t,\mathbf{x}) = F_1^\varepsilon(t,\mathbf{x})$$

where

$$F_1^\varepsilon(t,\mathbf{x}) = \frac{1}{\varepsilon}\sum_{r=1}^{R} \boldsymbol{v}_q \cdot \nabla\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{v}_r) - u^\varepsilon(t,\mathbf{x})\right]. \tag{11}$$

Let P_t^ε be the semigroup associated with $\mathbf{X}^\varepsilon(t)$, so that

$$P_t^\varepsilon f(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{X}^\varepsilon(t))\,|X^\varepsilon(0)=\mathbf{x}\right].$$

Clearly, if f is bounded, then $|P_t^\varepsilon f(\mathbf{x})| \le \|f\|_\infty$. In particular, $|u^\varepsilon(t,\mathbf{x})|$ is bounded uniformly with respect to ε. It is straightforward to check that we can write the solution $A_q^\varepsilon(t,\mathbf{x})$ as

$$A_q^\varepsilon(t,\mathbf{x}) = P_t\left[\boldsymbol{v}_q\cdot\nabla g\right](\mathbf{x}) + \int_0^t P_{t-s}\,F_1^\varepsilon(s,\mathbf{x})\,\mathrm{d}s,$$

where g is the fixed observable used as the initial condition in (8). Thus,

$$|A_q^\varepsilon(t,\mathbf{x})| \le \|\boldsymbol{v}_q\cdot\nabla g\|_\infty + \int_0^t \|F_1^\varepsilon(s,\cdot)\|_\infty\,\mathrm{d}s. \tag{12}$$

Since $u^\varepsilon(t,\cdot)$ is C^1, we have

$$\frac{1}{\varepsilon}\left[u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{v}_r) - u^\varepsilon(t,\mathbf{x})\right] = \int_0^1 A_r^\varepsilon(t,\mathbf{x}+w\,\varepsilon\,\boldsymbol{v}_r)\,w\,\mathrm{d}w.$$

Substituting into (11), we obtain

$$\|F_1^\varepsilon(t,\cdot)\|_\infty \le C\sum_{r=1}^{R}\|\widetilde{\lambda}_r^\varepsilon\|_{C^1}\|A_r^\varepsilon(t,\cdot)\|_\infty\,\mathrm{d}s.$$

Therefore, using (12), we get

$$\max_q\|A_q^\varepsilon(t,\cdot)\|_\infty \le \max_q\|A_q^\varepsilon(0,\cdot)\|_\infty + K_1\int_0^T \max_q\|A_q^\varepsilon(s,\cdot)\|_\infty\,\mathrm{d}s,$$

so that

$$\max_q\|A_q^\varepsilon(t,\cdot)\|_\infty \le C_1 e^{K_1 T}.$$

Similarly,

$$\partial_t B_{p,q}^\varepsilon(t,\mathbf{x}) - \mathcal{L}^\varepsilon B_{p,q}^\varepsilon(t,\mathbf{x}) = F_2^\varepsilon(t,\mathbf{x}),$$

where

$$\begin{aligned}F_2^\varepsilon(t,\mathbf{x}) = {}& \frac{1}{\varepsilon}\sum_{r=1}^{R}(\boldsymbol{v}_p\otimes\boldsymbol{v}_q) : \nabla\nabla\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{v}_r) - u^\varepsilon(t,\mathbf{x})\right]\\ &+ \frac{1}{\varepsilon}\sum_{r=1}^{R}\boldsymbol{v}_q\cdot\nabla\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[A_p^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{v}_r) - A_p^\varepsilon(t,\mathbf{x})\right]\\ &+ \frac{1}{\varepsilon}\sum_{r=1}^{R}\boldsymbol{v}_p\cdot\nabla\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\left[A_q^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{v}_r) - A_q^\varepsilon(t,\mathbf{x})\right].\end{aligned}$$

Thus it follows that

$$B^{\varepsilon}_{p,q}(t,\mathbf{x}) = P_t\left[(\boldsymbol{\nu}_p\otimes\boldsymbol{\nu}_q):\nabla\nabla g\right](\mathbf{x}) + \int_0^t P_{t-s}\,F^{\varepsilon}_2(s,\mathbf{x})\,\mathrm{d}s,$$

where

$$\|F^{\varepsilon}_2(t,\cdot)\|_{\infty} \le C\sum_{r=1}^{R}\|\widetilde{\lambda}^{\varepsilon}_r\|_{C^2(\mathbb{R}^N)}\left(||A_r(t,\cdot)||_{\infty} + ||B_{r,q}(t,\cdot)||_{\infty} + ||B_{r,p}(t,\cdot)||_{\infty}\right).$$

It follows that

$$\max_{p,q}\left\|B^{\varepsilon}_{p,q}(t,\cdot)\right\|_{\infty} \le \max_{p,q}\left\|B^{\varepsilon}_{p,q}(0,\cdot)\right\|_{\infty} + K_2\int_0^T\left(\max_{p,q}\left\|B^{\varepsilon}_{p,q}(s,\cdot)\right\|_{\infty} + \max_{p}\left\|A_p(s,\cdot)\right\|_{\infty}\right)\mathrm{d}s,$$

which implies the second inequality in (10) by Gronwall's inequality. The proof of the third inequality in (10) follows analogously. □

Assumption 3.3. We assume that the blending functions β_r, $r = 1,2,\ldots,R$ satisfy $\beta_r \in C^0(\mathbb{R}^N)$ and $\beta_r(\mathbf{x}) \in [0,1]$, for all $\mathbf{x}\in\mathbb{R}^N$.

Given the extended propensities, we can apply the same large-volume rescaling to the hybrid process (4) to obtain a jump-diffusion $\mathbf{Z}^{\varepsilon}(t)$ given by

$$\begin{aligned}\mathbf{Z}^{\varepsilon}(t) = \mathbf{Z}^{\varepsilon}(0) &+ \varepsilon\sum_{r=1}^{R}P_r\left(\frac{1}{\varepsilon}\int_0^t \beta_r(\mathbf{Z}^{\varepsilon}(s))\,\widetilde{\lambda}^{\varepsilon}_r(\llbracket\mathbf{Z}^{\varepsilon}(s)\rrbracket_{\varepsilon})\,\mathrm{d}s\right)\boldsymbol{\nu}_r\\ &+\sum_{r=1}^{R}\boldsymbol{\nu}_r\int_0^t(1-\beta_r(\mathbf{Z}^{\varepsilon}(s))\,\widetilde{\lambda}^{\varepsilon}_r(\mathbf{Z}^{\varepsilon}(s))\,\mathrm{d}s\\ &+\sum_{r=1}^{R}\sqrt{\varepsilon}\,\boldsymbol{\nu}_r\int_0^t\sqrt{(1-\beta_r(\mathbf{Z}^{\varepsilon}(s))\,\widetilde{\lambda}^{\varepsilon}_r(\mathbf{Z}^{\varepsilon}(s))}\,\mathrm{d}W_r(s),\end{aligned}$$

where $\{W_r\}_{r=1}^R$ and $\{P_r\}_{r=1}^R$ are standard Wiener and Poisson processes, respectively, all mutually independent, and $\llbracket\mathbf{x}\rrbracket_{\varepsilon}$ is the closest in $\varepsilon\mathbb{L}$ to $\mathbf{x}$, or equivalently $\llbracket\mathbf{x}\rrbracket_{\varepsilon} = \varepsilon\llbracket\mathbf{x}/\varepsilon\rrbracket$. The generator of this process is given by

$$\begin{aligned}\mathcal{G}^{\varepsilon}f(\mathbf{x}) = &\frac{1}{\varepsilon}\sum_{r=1}^{R}\beta_r(\mathbf{x})\,\widetilde{\lambda}^{\varepsilon}_r(\llbracket\mathbf{x}\rrbracket_{\varepsilon})\left[f(\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)-f(\mathbf{x})\right]\\ &+\sum_{r=1}^{R}(1-\beta_r(\mathbf{x}))\,\widetilde{\lambda}^{\varepsilon}_r(\mathbf{x})\,\boldsymbol{\nu}_r\cdot\nabla f(\mathbf{x})\\ &+\frac{\varepsilon}{2}\sum_{r=1}^{R}(1-\beta_r(\mathbf{x}))\,\widetilde{\lambda}^{\varepsilon}_r(\mathbf{x})\,(\boldsymbol{\nu}_r\otimes\boldsymbol{\nu}_r):\nabla\nabla f(\mathbf{x}).\end{aligned}$$

We now obtain a weak error estimate between the processes $\mathbf{X}^{\varepsilon}(t)$ and $\mathbf{Z}^{\varepsilon}(t)$ in the large volume limit as $\varepsilon\to 0$.

Theorem 3.5. *Let blending functions satisfy Assumption 3.3. Let $g\in C^3(\mathbb{R}^N)$, then there exists a constant $C>0$, independent of ε, such that*

$$\left|\mathbb{E}_{\mathbf{x}}\left[g(\mathbf{X}^{\varepsilon}(t))\right]-\mathbb{E}_{\mathbf{x}}\left[g(\mathbf{Z}^{\varepsilon}(t))\right]\right| \le C\varepsilon^2, \quad \text{for } t\in[0,T], \tag{13}$$

where $\mathbf{X}^{\varepsilon}(0)=\mathbf{Z}^{\varepsilon}(0)=\mathbf{x}\in\varepsilon\,\Omega_0$.

Proof. Let $u^{\varepsilon}(t,\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[g(\mathbf{X}^{\varepsilon}(t))]$ and $v^{\varepsilon}(t,\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[g(\mathbf{Z}^{\varepsilon}(t))]$. We wish to obtain a bound on $E^{\varepsilon}(t,\mathbf{x}) = u^{\varepsilon}(t,\mathbf{x}) - v^{\varepsilon}(t,\mathbf{x})$. Then taking the derivative with respect to t and using the fact that $\llbracket\mathbf{x}\rrbracket_{\varepsilon}=\mathbf{x}$ for all $\mathbf{x}\in\varepsilon\Omega_0$, we obtain

$$\begin{aligned}\partial E^{\varepsilon}(t,\mathbf{x}) &= \partial_t u^{\varepsilon}(t,\mathbf{x}) - \partial_t v^{\varepsilon}(t,\mathbf{x}) = \mathcal{L}^{\varepsilon}u^{\varepsilon}(t,\mathbf{x}) - \mathcal{G}^{\varepsilon}v^{\varepsilon}(t,\mathbf{x})\\ &= \mathcal{L}^{\varepsilon}u^{\varepsilon}(t,\mathbf{x}) - \mathcal{G}^{\varepsilon}u^{\varepsilon}(t,\mathbf{x}) - (\mathcal{G}^{\varepsilon}u^{\varepsilon}(t,\mathbf{x}) + \mathcal{G}^{\varepsilon}v^{\varepsilon}(t,\mathbf{x}))\end{aligned}$$

implying that

$$\begin{aligned}\partial_t E^\varepsilon(t,\mathbf{x}) - \mathcal{G}^\varepsilon E^\varepsilon(t,\mathbf{x}) &= \left(\mathcal{L}^\varepsilon - \mathcal{G}^\varepsilon\right) u^\varepsilon(t,\mathbf{x}) \\ &= \sum_{r=1}^{R} (1-\beta_r(\mathbf{x}))\,\widetilde{\lambda}_r^\varepsilon(\mathbf{x}) \Big[\frac{u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r) - u^\varepsilon(t,\mathbf{x})}{\varepsilon} - \boldsymbol{\nu}_r\cdot\nabla u^\varepsilon(t,\mathbf{x}) \\ &\quad - \frac{\varepsilon}{2}(\boldsymbol{\nu}_r\otimes\boldsymbol{\nu}_r) : \nabla\nabla u^\varepsilon(t,\mathbf{x})\Big].\end{aligned}$$

Since $u^\varepsilon(t,\cdot)$ is in $C^3(\mathbb{R}^N)$, we can apply Taylor's theorem up to the second order on $u^\varepsilon(t,\mathbf{x}+\varepsilon\boldsymbol{\nu}_r)$ to obtain

$$\partial_t E^\varepsilon(t,\mathbf{x}) - \mathcal{G}^\varepsilon E^\varepsilon(t,\mathbf{x}) = c^\varepsilon(t,\mathbf{x}), \tag{14}$$

where

$$c^\varepsilon(t,\mathbf{x}) = \sum_{r=1}^{R} \frac{\varepsilon^2}{6}(1-\beta_r(\mathbf{x}))\,\widetilde{\lambda}_r^\varepsilon(\mathbf{x})\,(\boldsymbol{\nu}_r\otimes\boldsymbol{\nu}_r\otimes\boldsymbol{\nu}_r) : \nabla\nabla\nabla u^\varepsilon(t,\xi_r),$$

for some ξ_r lying on the line between $\mathbf{x}$ and $\mathbf{x}+\varepsilon\boldsymbol{\nu}_r$. From (14) and the fact that $E(0,\mathbf{x})=0$, it follows that

$$E^\varepsilon(t,\mathbf{x}) = \int_0^t R^\varepsilon_{t-s}\, c^\varepsilon(s,\mathbf{x})\,\mathrm{d}s,$$

where R^ε_t is the semigroup operator corresponding to $\mathbf{Z}^\varepsilon(t)$. Applying the uniform bound (10) we thus have that

$$\left|E^\varepsilon(t,\mathbf{x})\right| \le C\varepsilon^2 \int_0^t \sum_{r=1}^{R} \|C^\varepsilon_{r,r,r}(s,\cdot)\|_\infty \,\mathrm{d}s \le C_1\varepsilon^2 T e^{K_1 T}, \quad \text{for } t\in[0,T]. \qquad \square$$

Remark 3.6. The remainder term $c^\varepsilon(t,\mathbf{x})$ in equation (14) characterises the local error at the point $\mathbf{x}$. We note that it is non-zero only in regions where $\beta_r(\mathbf{x})\neq 1$. Intuitively we would expect this to imply that $\mathbf{Z}^\varepsilon(t)$ is a superior approximation to the standard CLE. However, the global error estimate we derived is too coarse to capture the distinction between the two diffusion approximations, and thus we have only shown that the two approximations are consistent: in that the hybrid scheme does no worse than the CLE in the large-volume limit.

4. Simulating the hybrid model

Equation (4) provides a general framework for simulating chemical systems which can capture both the discrete and continuum nature of a biochemical system. Any numerical scheme which can generate realisations of a jump-diffusion process with inhomogeneous jump rates with deterministic jump sizes can be used to simulate (4). In order to connect with numerical schemes previously used in the literature for the pure jump case we describe three possible numerical schemes for simulating the jump-diffusion dynamics. The first algorithm is a straightforward generalisation of the Gillespie SSA algorithm. The second algorithm is an adaptation of the Gibson–Bruck next reaction method [24], based on the generalised approach of [1]. The main distinction between the standard algorithms for simulating stochastic reaction networks and the current hybrid implementation is that for the latter case propensity functions will not remain constant between two consecutive reactions. Finally, the third algorithm provides an alternative approach based on Poisson thinning [45], which can be computationally more efficient provided that tight bounds on the propensity functions are available a priori.

For the sake of clarity, given the propensities $\lambda_1,\lambda_2,\ldots,\lambda_R$ and blending functions $\beta_1,\beta_2,\ldots,\beta_R$ define

$$\lambda'_j(\mathbf{x}) = \beta_j(\mathbf{x})\lambda_j(\llbracket\mathbf{x}\rrbracket), \quad \text{and} \quad \lambda''_j(\mathbf{x}) = (1-\beta_j(\mathbf{x}))\lambda_j(\mathbf{x}). \tag{15}$$

Pseudocodes of each approach are given as Algorithms 1, 2 and 3. They all have the same input, namely propensities $\lambda_1,\lambda_2,\ldots,\lambda_R$, blending functions $\beta_1,\beta_2,\ldots,\beta_R$, the stoichiometric matrix $\underline{\boldsymbol{\nu}} = (\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_R)$, the final time of simulation T, time steps for the CLE Δt and δt (here, $\delta t \ge \Delta t$) and initial state $\mathbf{X}(0)\in\mathbb{N}^N$.

4.1. Hybrid simulations based on the Gillespie SSA

The steps to simulate the jump-diffusion process (4) based on an extension of the Gillespie SSA are described in Algorithm 1. As we can see in regions where $\beta_r(x)$ are 1, the scheme reduces to the standard Gillespie SSA, and thus simulates the discrete dynamics exactly. Analogously, in regions where β_r are all zero one can use a larger time-step δt to evolve

Set $t = 0$.
while $t < T$ **do**
  if $\max_j \beta_j(\mathbf{X}(t)) = 0$ **then**
    Simulate CLE (3) up to time $t + \delta t$.
    Set $t = t + \delta t$.
  else if $\min_j \beta_j(\mathbf{X}(t)) = 1$ **then**
    Compute $\lambda_0 = \sum_{j=1}^{R} \lambda'_j(\mathbf{X}(t))$.
    Sample $\tau \sim -\log(u)/\lambda_0$ where $u \sim U[0, 1]$.
    Choose the next reaction r with probability $\lambda'_r(\mathbf{X}(t))/\lambda_0$.
    Set $\mathbf{X}(t + \tau) = \mathbf{X}(t) + \boldsymbol{\nu}_r$.
    Set $t = t + \tau$.
  else
    Compute $\lambda'_0 = \sum_{j=1}^{R} \lambda'_j(\mathbf{X}(t))$.
    Sample $\tau \sim -\log u/\lambda'_0$, where $u \sim U[0, 1]$.
    Choose the next reaction r with probability $\lambda'_r(\mathbf{X}(t))/\lambda'_0$.
    if $\tau < \Delta t$ **then**
      Simulate CLE (16) up to time $t + \tau$ and set $\mathbf{X}(t + \tau) = \mathbf{X}(t + \tau) + \boldsymbol{\nu}_r$.
      Set $t = t + \tau$.
    else
      Simulate CLE (16) up to time $t + \Delta t$.
      Set $t = t + \Delta t$.
    end
  end
end

Algorithm 1: Generating approximate realisations of hybrid model (4).

CLE (3) since it is not necessary to approximate the solutions of (4) in such regions, which can only be done with $\mathcal{O}(\Delta t)$ accuracy. In the intermediate regime for the continuous part of the dynamics the following CLE is used

$$\mathrm{d}\mathbf{X}(t) = \sum_{j=1}^{R} \lambda''_j(\mathbf{X}(s))\, \boldsymbol{\nu}_j\, \mathrm{d}t + \boldsymbol{\nu}_j \sqrt{\lambda''_j(\mathbf{X}(s))}\, \mathrm{d}W_j(s). \tag{16}$$

If during the CLE time step $[t, t + \Delta t)$ a discrete event is occurring at time $t + \tau$ we simulate the CLE up to that time and then add the discrete event. To simulate the diffusion part of the hybrid scheme we make use of the weak trapezoidal method described in [3]. Given the current state $\mathbf{X}^n$ we perform the following two steps to obtain $\mathbf{X}^{n+1}$:

$$\mathbf{X}^* = \mathbf{X}^n + \frac{\Delta t}{2} \sum_{j=1}^{R} \boldsymbol{\nu}_j \lambda''_j\left(\mathbf{X}^n\right) + \sqrt{\frac{\Delta t}{2}} \sum_{j=1}^{R} \boldsymbol{\nu}_j \lambda''_j(\mathbf{X}^n)\, \xi_j,$$

$$\mathbf{X}^{n+1} = \mathbf{X}^* + \frac{\Delta t}{2} \sum_{j=1}^{R} h_j(\mathbf{X}^*, \mathbf{X}^n) + \sqrt{\frac{\Delta t}{2}} \sum_{j=1}^{R} \sqrt{\left[h_j(\mathbf{X}^*, \mathbf{X}^n)\right]^+}\, \boldsymbol{\nu}_j\, \xi'_j,$$

where $h_j(\mathbf{X}^*, \mathbf{X}^n) = 2\lambda''_j(\mathbf{X}^*) - \lambda''_j(\mathbf{X}^n)$, $[a]^+ = \max(0, a)$ and $\xi_1, \xi_2, \ldots, \xi_R$, $\xi'_1, \xi'_2, \ldots \xi'_R$ are mutually independent standard Gaussian random variables.

4.2. Hybrid simulations based on the next reaction method

A second algorithm, based on [1], for simulating (4) is described in Algorithm 2. While it is entirely equivalent to the standard Next Reaction Method, the modified scheme keeps explicit track of the internal times T_k and the next firing time F_k of each Poisson process P_k which simplifies integrating diffusion steps into the scheme. We note that the hybrid scheme described in [22] also employs a similar discretisation. Again in the presence of diffusion, it is no longer true that the propensity $\lambda'_j(\mathbf{X}(s))$ is constant from t until the next reaction. Computing the next reaction time is equivalent to solving the following first passage time problem:

Compute $\inf\{s \geq t : T_r(s) = F_r(t) - \log u \text{ for some } r = 1, 2, \ldots, R\}$ where:

$$\begin{aligned} \mathrm{d}\mathbf{X}(s) &= \sum_{j=1}^{R} \lambda''_j(\mathbf{X}(s))\, \boldsymbol{\nu}_j\, \mathrm{d}t + \boldsymbol{\nu}_j \sqrt{\lambda''_j(\mathbf{X}(s))}\, \mathrm{d}W_j(s) \\ \mathrm{d}T_r(s) &= \lambda'_r(\mathbf{X}(s))\, \mathrm{d}s, \quad r = 1, 2, \ldots, R. \end{aligned} \tag{17}$$

In Algorithm 2, we use an Euler discretisation of $T_r(s)$. While Algorithm 2 is more complex than the previous scheme, it has numerous advantages. Primarily, its implementation can be significantly optimised, in particular, the next reaction can

Generate R independent, $U[0,1]$ random numbers u_j, $j = 1, 2, \ldots, R$.
Set $t = 0$. Set $F_j = -\log(u_j)$ and $T_j = 0$, for each $j = 1, 2, \ldots, R$.
Compute the weighted propensities $\lambda'_j = \beta(\mathbf{X}(0))\lambda_j(\mathbf{X}(0))$.
while $t < T$ **do**
  if $\max_j \beta_j(\mathbf{X}(t)) = 0$ **then**
    Simulate CLE (3) up to time $t + \delta t$. Set $\tau = \delta t$.
  else
    Set $\tau_j = (F_j - T_j)/\lambda'_j$ for $j = 1, 2, \ldots, R$.
    Let $r = \text{argmin}_j\{\tau_j\}$ and set $\tau = \tau_r$.
    if $\min_j \beta_j(\mathbf{X}(t)) = 1$ **then**
      Set $\mathbf{X}(t + \tau) = \mathbf{X}(t) + \boldsymbol{\nu}_r$.
      Sample $u \sim U[0,1]$ and set $F_r = F_r - \log(u)$.
    else
      if $\tau < \Delta t$ **then**
        Simulate CLE (16) up to time $t + \tau$.
        Set $\mathbf{X}(t + \tau) = \mathbf{X}(t) + \boldsymbol{\nu}_r$.
        Sample $u \sim U[0,1]$ and set $F_r = F_r - \log(u)$.
      else
        Simulate CLE (16) up to time $t + \Delta t$. Set $\tau = \Delta t$.
      end
    end
    Set $t = t + \tau$. Set $T_j = T_j + \lambda'_j \tau$ for $j = 1, 2, \ldots, R$.
    Update the propensities $\lambda'_j = \beta_j(\mathbf{X}(t))\lambda_j([\![\mathbf{X}(t)]\!])$.
  end
end

Algorithm 2: Generating approximate realisations of hybrid model (4).

be computed efficiently using an indexed priority queue. Moreover, as noted in [1] this method can be further generalised to permit propensities which are time-dependent or involve delays.

4.3. Thinning method

The main drawback of the previous two approaches is the necessity to solve the first passage time problem (17) with sufficient accuracy. As in the previous algorithm, solving (17) numerically will introduce error arising from the piecewise constant approximation of integral $T_r(t)$. As noted in [58], in many applications this constrains the maximum step-size which can be used to simulate the next reaction time, which can give rise to a significant increase in computational cost to simulate the model.

In the special case where one can bound the value of the weighted propensities $\{\lambda'_r\}_{r=1}^R$, it is possible to use a third method, based on standard thinning methods for sampling inhomogeneous Poisson processes, see for example [47,45] and more recently [55,58]. While one can find such bounds for many chemical systems, the added caveat is that the bounds must be known a priori, and choosing them too loosely will severely degrade the performance of the scheme.

To this end, we shall assume that there exist constants $\Lambda_1, \Lambda_2, \ldots, \Lambda_R$, where

$$\lambda'_r(\mathbf{x}) \leq \Lambda_r, \quad \text{for all } \mathbf{x}. \tag{18}$$

These constants will form the additional input of Algorithm 3. Suppose that $\beta_r(\mathbf{X}(t)) > 0$ for some r. To compute the next jump time of the r-th reaction, we sample from a dominating homogeneous process with rate Λ_r, so that the next jump time for the r-th reaction occurs at time $t + \tau_r$ where

$$\tau_r = -\frac{\log(u)}{\Lambda_r}, \quad u \sim U[0,1].$$

Suppose that $t + \tau_r$ is the first jump occurring after the current time t. To determine whether the reaction r will occur at time $t + \tau_r$, we sample $u' \sim U[0,1]$ and perform the reaction only if

$$\Lambda_r u' \leq \lambda'_r(\mathbf{X}(t + \tau_r)), \tag{19}$$

where $\mathbf{X}(t + \tau_r)$ is the state of the process after simulating the Langevin dynamics from time t to $t + \tau_r$. This thinning approach can be integrated into the Gillespie SSA, demonstrated in Algorithm 3. At each timestep, three cases can occur. If $\min \beta_r(\mathbf{X}(t)) = 1$ (i.e. the process $\mathbf{X}(t)$ is a pure jump process) we use the standard Gillespie SSA. If $\max \beta_r(\mathbf{X}(t)) = 0$ (i.e. the process is purely diffusive), then we perform a "macro-step" of the CLE dynamics of size δt. The final case is where there is both diffusion and jumps, we simulate the homogeneous dominating process with rate $\Lambda_0 = \sum_{r=1}^R \Lambda_r$, and accept/reject according to condition (19).

The main caveat of this approach is the necessity to know *a priori* the upper bounds Λ_r, assuming such bounds exist. Care must be taken so that the bounds are not too pessimistic, otherwise the dominating homogeneous Poisson process

Set $t = 0$.
while $t < T$ **do**
  if $\max_{r=1,2,\ldots,R} \beta_r(\mathbf{X}(t)) = 0$ **then**
    Simulate CLE (3) up to time $t + \delta t$.
    Set $t = t + \delta t$.
  else if $\min_{r=1,2,\ldots,R} \beta_r(\mathbf{X}(t)) = 1$ **then**
    Let $\lambda_0 = \sum_{r=1}^{R} \lambda'_r(\mathbf{X}(t))$.
    Let $u \sim U[0,1]$ and $\tau = -\log(u)/\lambda_0$.
    Choose index r with probability $\lambda'_r(\mathbf{X}(t))/\lambda_0$.
    Set $\mathbf{X}(t+\tau) = \mathbf{X}(t) + \boldsymbol{\nu}_r$.
    Set $t = t + \tau$.
  else
    Let $\Lambda_0 = \sum_{r=1}^{R} \Lambda_r$.
    Let $u \sim U[0,1]$ and set $\tau = -\log(u)/\Lambda_0$.
    Simulate CLE (16) up to time $t + \tau$ using stepsize Δt.
    Let $u' \sim U[0,1]$.
    if $\Lambda_0 u' \leq \sum_{r=1}^{R} \lambda'_r(\mathbf{X}(t+\tau))$ **then**
      Let r be smallest index such that $\Lambda_0 u' \leq \sum_{j=1}^{r} \lambda'_j(\mathbf{X}(t+\tau))$.
      Set $\mathbf{X}(t+\tau) = \mathbf{X}(t) + \boldsymbol{\nu}_r$.
      Set $t = t + \tau$.
    end
end

Algorithm 3: Generating approximate realisations of hybrid model (4).

will fire very rapidly when the system lies within a blending region. In such cases the bounds can be tuned by running exploratory simulations and keeping track of the acceptance rate for each reaction.

5. Numerical investigations

We illustrate the main features of the hybrid framework described in Section 2.2 and demonstrate the use of Algorithms 1, 2 and 3 by considering three numerical examples. Each of these examples were implemented in the programming language Julia [8].

5.1. Lotka–Volterra model

As a first example, we consider a chemical system consisting of two reacting chemical species A and B undergoing the following reactions

$$A \xrightarrow{k_1} 2A, \qquad A + B \xrightarrow{k_2} 2B, \qquad B \xrightarrow{k_3} \emptyset. \tag{20}$$

The chemicals A and B can be considered to be in a "predator–prey" relationship with A and B as prey and predator, respectively. The reaction-rate equations corresponding to reactions (20) would then be the standard Lotka–Volterra model. We choose the dimensionless parameters $k_1 = 2.0$, $k_2 = 0.002$ and $k_3 = 2.0$. The initial condition is chosen to be $A(0) = 50$ and $B(0) = 60$. A histogram generated from 10^3 independent SSA simulations of this system up to time $T = 5$ is shown in Fig. 2. The dashed line depicts the evolution of the deterministic reaction rate equation starting from the same initial point. One sees that the nonequilibrium dynamics force the system to spend time in both low and high concentration regimes. Due to the time spent in states with high propensity, the SSA is computationally expensive to simulate. It is clear that away from the boundary, using an approximation such as the CLE would be computationally beneficial. The CLE corresponding to (20), choosing the multiplicative noise as described in [27], is given by

$$\begin{aligned}
\mathrm{d}A(t) &= \Big(k_1 A(t) - k_2 A(t) B(t)\Big)\,\mathrm{d}t + \sqrt{k_1 A(t)}\,\mathrm{d}W_1(t) - \sqrt{k_2 A(t) B(t)}\,\mathrm{d}W_2(t), \\
\mathrm{d}B(t) &= \Big(k_2 A(t) B(t) - k_3 B(t)\Big)\,\mathrm{d}t - \sqrt{k_2 A(t) B(t)}\,\mathrm{d}W_2(t) - \sqrt{k_3 B(t)}\,\mathrm{d}W_3(t),
\end{aligned}$$

where $W_1(t)$, $W_2(t)$ and $W_3(t)$, are three standard independent Brownian motions. For a non-negative initial condition, the process $(A(t), B(t))$ will remain nonnegative, however, this will not be the case for fixed-timestep discretisation. In particular, an Euler discretisation (A_n, B_n) will contain a term of the form $-\sqrt{k_3 B_n \Delta t}\,\xi$, where ξ is a standard Gaussian random variable, which can cause the discretised process to cross the $B = 0$ axis if the process sufficiently close to this line. Thus, it is essential that reflective boundary conditions are imposed to ensure positivity. However, even if positivity is guaranteed, there is no reason to believe that the CLE will correctly approximate the dynamics near the axes. This motivates the use of the hybrid model to efficiently simulate this chemical system.

To simulate the hybrid model we use Algorithm 1, choosing blending functions β_1, β_2 and β_3 as described in (6). We simulate the Langevin dynamics in the blending region using a timestep of size $\Delta t = 10^{-3}$, and a timestep $\delta t = 10^{-2}$ outside

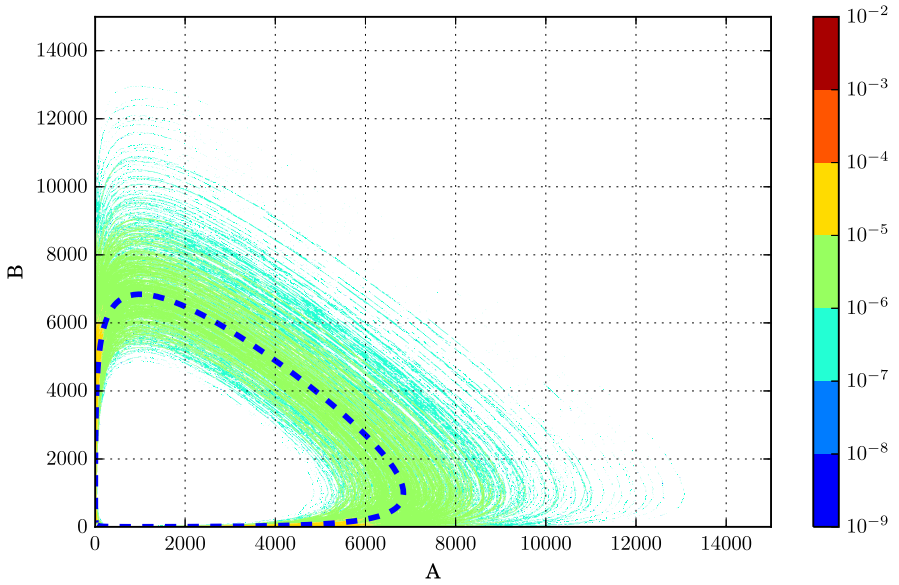


Fig. 2. Histogram of 10^3 SSA simulations of (20) up to time $T = 5.0$ starting from $A(0) = 50$ and $B(0) = 60$. The dashed line is the solution of the corresponding deterministic reaction rate equations.

the blending regions. In Fig. 3(a), we plot $\mathbb{E}(A(t))$ at times $T = 1, 2, \ldots, 6$, generated from 10^4 independent realisations of each model. Error bars denote 95% confidence intervals. The hybrid models were simulated for different values of I_1^i and I_2^i, $i = 1, 2, 3$, however, the results were not plotted as the Monte Carlo error was too large to distinguish between the schemes. The hybrid scheme displayed in Fig. 3(a) has blending regions with parameters $I_1^i = 25.0$ and $I_2^i = 35.0$. While all models agree approximately at small times, for larger T the averages generated from SSA and CLE differ significantly. Indeed, for $T = 5$ (corresponding to a single period of the deterministic system) the means from the SSA and CLE differ by three orders of magnitude. On the other hand, the hybrid scheme remains in good agreement with the SSA. In Fig. 3(b) we compare the average computational (CPU) time to simulate each model up to time T, averaged over 10^4 realisations. This was measured in seconds, using the standard Julia functions tic and toc. To study qualitatively the effect of the choice of blending regions, the computational cost of the hybrid scheme was plotted for three different choices of blending regions, namely $(I_1^i, I_2^i) = (5, 15)$, $(10, 25)$ and $(25, 35)$. For small T the SSA and hybrid schemes require a comparable amount of computational effort. However, as T increases, the computational cost of the SSA scheme dramatically increases, while the cost of the hybrid scheme remains approximately constant. To compute the average value at time $T = 6$, the hybrid scheme was on average 2 orders of magnitude cheaper to run. It was observed that the simulations were very robust to the choice of blending function. Indeed, there was only negligible difference between the three choices in terms of error. As expected, the computational effort was smaller when the blending region is closer to the boundary. However, relative to the computational cost of the SSA and CLE, varying the blending region does not significantly alter performance.

For these simulations, the value of Δt and δt were chosen manually by computing the error for a number of short exploratory runs. A more sophisticated implementation of the hybrid model would require an adaptive scheme for the Langevin part of the process.

5.2. Steady state simulations

As a second example, we consider a chemical system consisting of two species A and B in a reactor of volume V. The species are subject to the following system of four chemical reactions [16]:

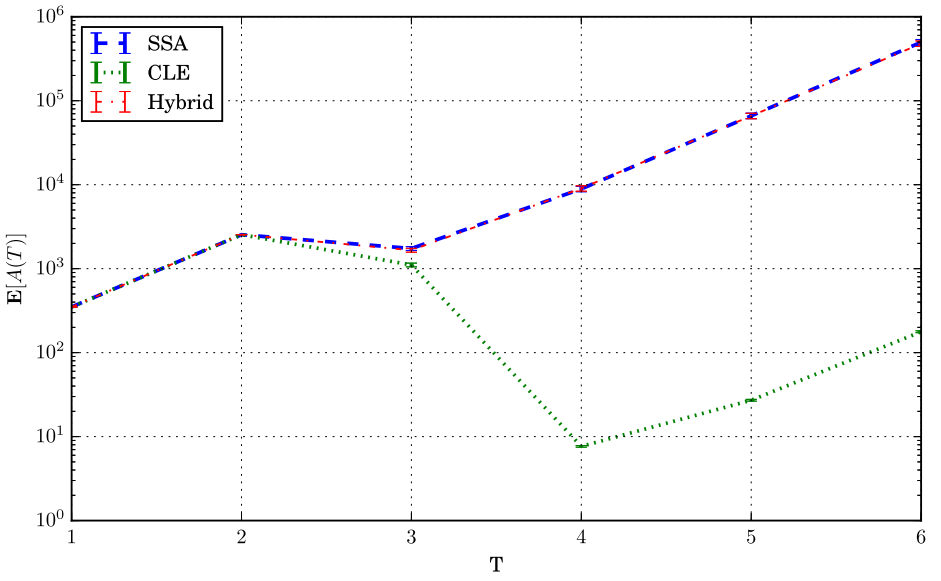
$$A + A \xrightarrow{k_1} \emptyset, \qquad A + B \xrightarrow{k_2} \emptyset, \qquad \emptyset \xrightarrow{k_3} A, \qquad \emptyset \xrightarrow{k_4} B. \tag{21}$$

This corresponds to a jump process $\mathbf{X}(t)$ having stoichiometric vectors

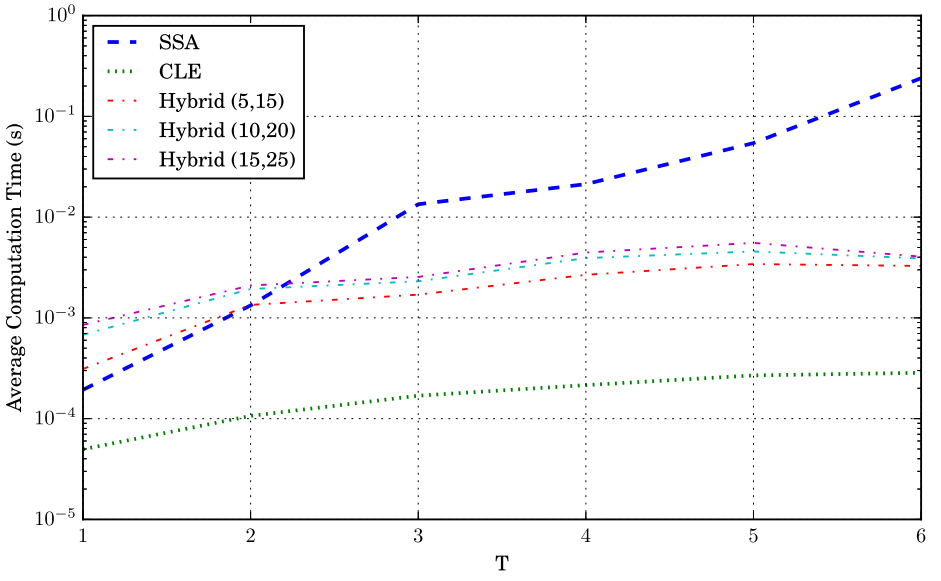
$$\boldsymbol{\nu}_1 = (-2, 0)^\top, \quad \boldsymbol{\nu}_2 = (-1, -1)^\top, \quad \boldsymbol{\nu}_3 = (1, 0)^\top, \quad \boldsymbol{\nu}_4 = (0, 1)^\top,$$

with corresponding propensities (depending on the volume V):

$$\lambda_1(a, b) = \frac{k_1 a(a-1)}{V}, \quad \lambda_2(a, b) = \frac{k_2 ab}{V}, \quad \lambda_3(a, b) = k_3 V, \quad \lambda_4(a, b) = k_4 V.$$



(a) Plot of $\mathbb{E}(A(T))$ as a function of T for the SSA, CLE and hybrid schemes



(b) Average computational (CPU) time in seconds as a function of T.

Fig. 3. Numerical simulations of the Lotka–Volterra model (20) using SSA, CLE and hybrid schemes.

The dimensionless reaction rates are given by $k_1 = 10^{-3}$, $k_2 = 10^{-2}$, $k_3 = 1.2$ and $k_4 = 1$. As a first numerical experiment, we compute the evolution of the distribution of (A, B) over time. We assume that $V = 0.25$, and that the initial distribution is a "discrete" Gaussian mixture, namely the Gaussian mixture

$$\rho_0 = \mathcal{N}((30, 10), 1) + \mathcal{N}((20, 30), 1), \tag{22}$$

restricted to the lattice $\mathbb{N}^2$. For each scheme (SSA, CLE and hybrid), the distribution is approximated by a histogram generated from 10^7 independent realisations of the process. The CLE was simulated using the weak second order trapezoidal scheme described in [3]. To ensure positivity of the CLE, reflective boundary conditions were imposed at the boundary of the orthant. The hybrid scheme was simulated the hybrid next reaction scheme detailed in Algorithm 2. The timestep was chosen to be $\Delta t = \delta t = 0.1$. The blending functions for the hybrid scheme were chosen according to (6), with $(I_1^i, I_2^i) = (5, 10)$, for $i = 1, \ldots, 4$. In Fig. 4 we plot the distribution approximated using each scheme at times $t = 0, 1, 10$ and 100. As ex-

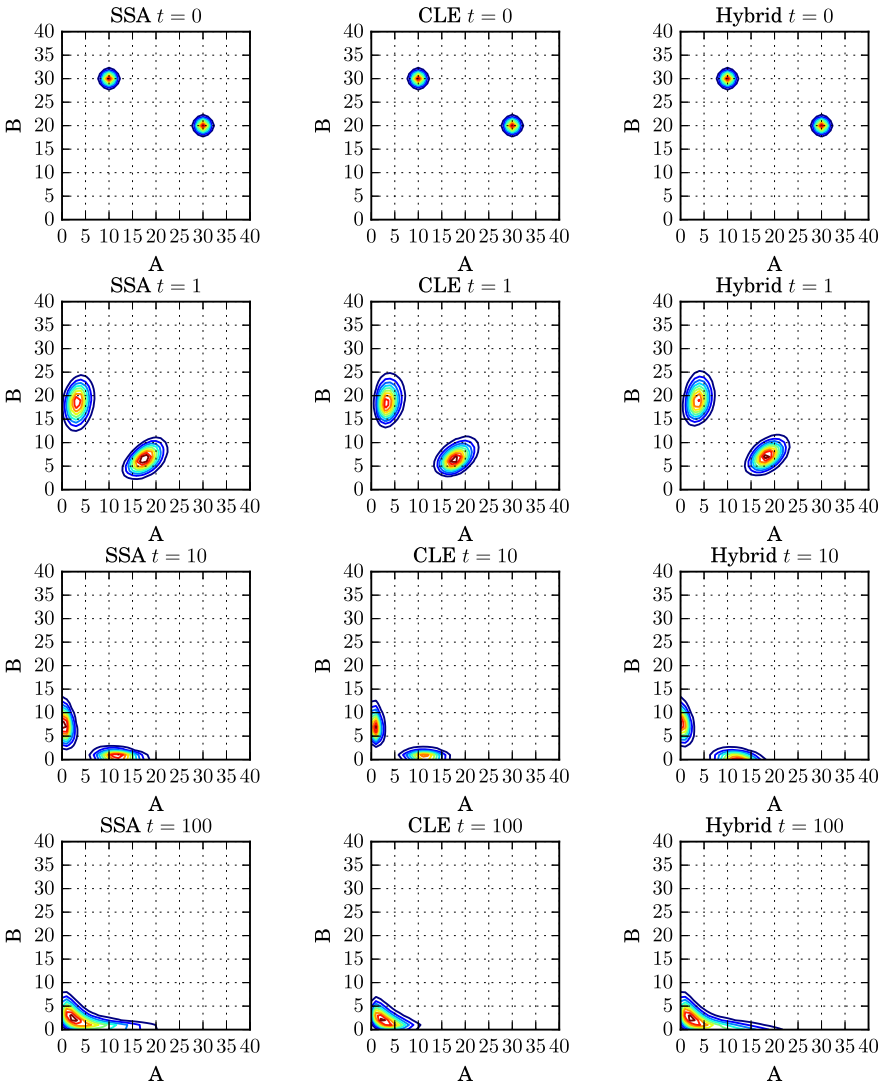


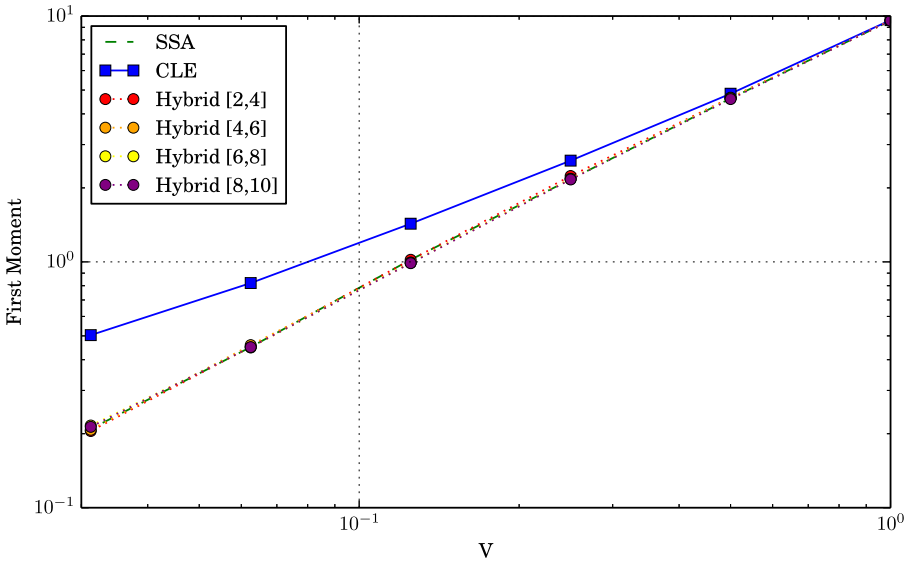
Fig. 4. Time evolution of the biochemical system (21) using SSA, CLE and the hybrid scheme starting from initial distribution ρ_0 given by (22).

pected, when the concentrations of A and B remain abundant, all three models agree. As the distribution approaches the low concentration regions, the discrete nature of the chemical system becomes important, and the CLE is no longer able to correctly capture the dynamics. Indeed, at time 100 one observes a significant difference between the SSA and CLE distributions. On the other hand, the hybrid scheme provides a good approximation to the SSA at all times, but benefiting from a computational advantage in the large concentration regimes.

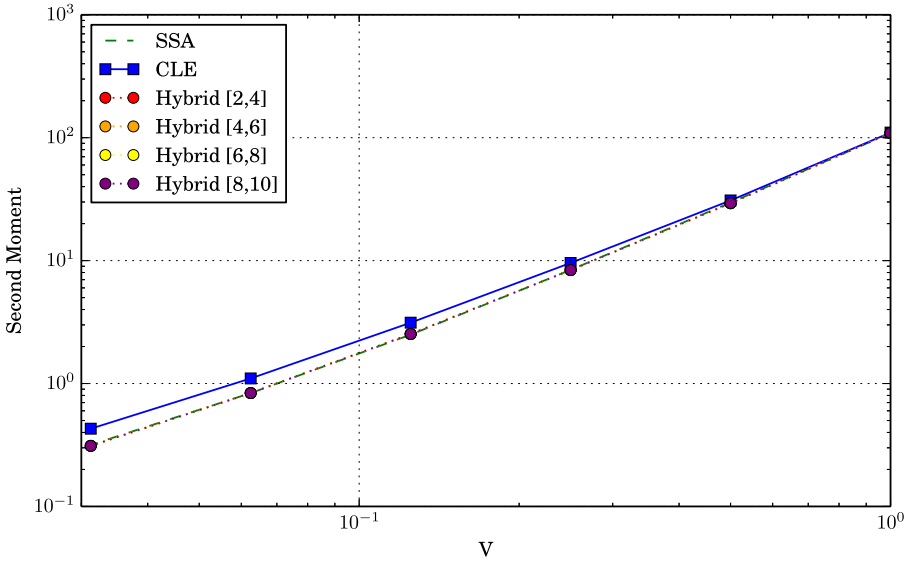
The corresponding Markov jump process $\mathbf{X}(t)$ can be shown to possess a unique stationary steady state [16]. We use all three models to compute the first two moments M_1 and M_2 of the stationary distribution, for decreasing values of V. The moments were approximated using ergodic average of the discretised schemes, i.e.

$$M_1 \approx \frac{1}{T}\sum_{t_i \leq T}(t_{i+1}-t_i)A_i \text{ and } M_2 \approx \frac{1}{T}\sum_{t_i \leq T}(t_{i+1}-t_i)\,(A_i)^2\,,$$

where (A_i, B_i) is the value of the discretised process at time t_i and $0 < t_1 < t_2 < \ldots < t_N = T$ are the jump times of the process. Each process was simulated up to $T = 5 \cdot 10^6$. For the hybrid and CLE schemes, a timestep of $\delta_t = \Delta t = 10^{-2}$ was used throughout. To demonstrate the robustness of the hybrid scheme with respect to the choice of blending region we plot the first two moments of the hybrid scheme for 4 different blending regions, namely $(I_1^i, I_2^i) = (2,4)$, $(4,6)$, $(6,8)$, $(8,10)$, respectively, for $i = 1, \ldots, 4$. The blending region was chosen as in the previous example. The first two moments are plotted in Fig. 5 for $V = 2^{-i}$ where $i = 0, 1, 2 \ldots, 6$. While there is good agreement between all three schemes for V large, the CLE



(a) *Comparison of the first moments for the three different schemes, i.e. SSA, CLE and hybrid with different choices of blending region.*



(b) Comparison of the second moments as a function of V.

Fig. 5. Numerical simulations of the biochemical system (21) using SSA, CLE and hybrid schemes.

consistently overestimates the moments when V is small. On the other hand, the hybrid scheme remains robust to this rescaling, regardless of the choice of blending region.

5.3. Exit time calculation for the birth–death problem

As a final example, we consider the problem of computing the mean extinction time (MET) for a one-dimensional birth–death process, namely a system of two reactions for one chemical species A:

$$\begin{aligned} &A \to \emptyset \text{ with propensity } \lambda_1(n) = k_1 n, \\ &\emptyset \to A \text{ with propensity } \lambda_2(n) = k_2 \text{ if } 0 < n < K, \text{ otherwise } 0. \end{aligned} \tag{23}$$

By assuming that $\lambda_2(n) = 0$ for all $n \geq K$, the state of the system lies within the finite domain $\{0, 1, 2, \ldots, K\}$. For $k_1 > 0$, the birth–death process will hit the extinction state $x = 0$ with probability 1. We denote by MET(n) the MET of the process starting from $A(0) = n$. Following directly the approach of [14, Section 2.1] (also see [23,56]), we obtain

$$\mathrm{MET}(n) = \frac{1}{k_1} \sum_{m=1}^{n} \sum_{j=0}^{K-m+1} \left(\frac{k_1}{k_2}\right)^j \frac{(m-1)!}{(j+m-1)!}, \qquad \text{for } n = 1, 2, \ldots, K. \tag{24}$$

The corresponding CLE is given by

$$\mathrm{d}Y(t) = (k_1 - k_2\, Y(t))\,\mathrm{d}t + \sqrt{k_1}\,\mathrm{d}W_1(t) - \sqrt{k_2\, Y(t)}\,\mathrm{d}W_2(t), \tag{25}$$

for standard independent Brownian motions $W_1(t)$ and $W_2(t)$. The mean first time of $Y(t)$ reaching 0 starting from $Y(0) = x$ can be calculated explicitly as

$$2 \int_0^x e^{-\Phi(y)} \int_y^K \frac{e^{\Phi(z)}}{k_2 + k_1 z}\, dz\, \mathrm{d}y, \tag{26}$$

where Φ is the potential

$$\Phi(x) = \frac{4k_2}{k_1} \log\left(1 + \frac{k_1 x}{k_2}\right) - 2x.$$

Since $n_e = k_2/k_1$ is a unique solution of $\lambda_1(n) = \lambda_2(n)$, the stochastic birth–death process will fluctuate around n_e for a long time before eventually going extinct. In general, we expect that any approximation which correctly describes the extinction time behaviour of the birth–death process must accurately capture the behaviour of the process particularly near $n = 0$ and $n = n_e$ (and possibly all points in between). If n_e is large, then a Gaussian approximation (e.g. CLE or the system size expansion) would accurately capture the fluctuations around the quasi-equilibrium. However, as observed in [14] such approximations would suffer close to $n = 0$. This suggests that the hybrid scheme with a blending region supported between $n = 0$ and $n = n_e$ would be a good candidate for an approximation to the process. Similar observations have been in more general chemical systems [35,30], where it is observed that diffusion approximations of jump processes are not able to correctly capture rare events, even when the system is in a regime where the CLE correctly captures both the transient and stationary dynamics of the process.

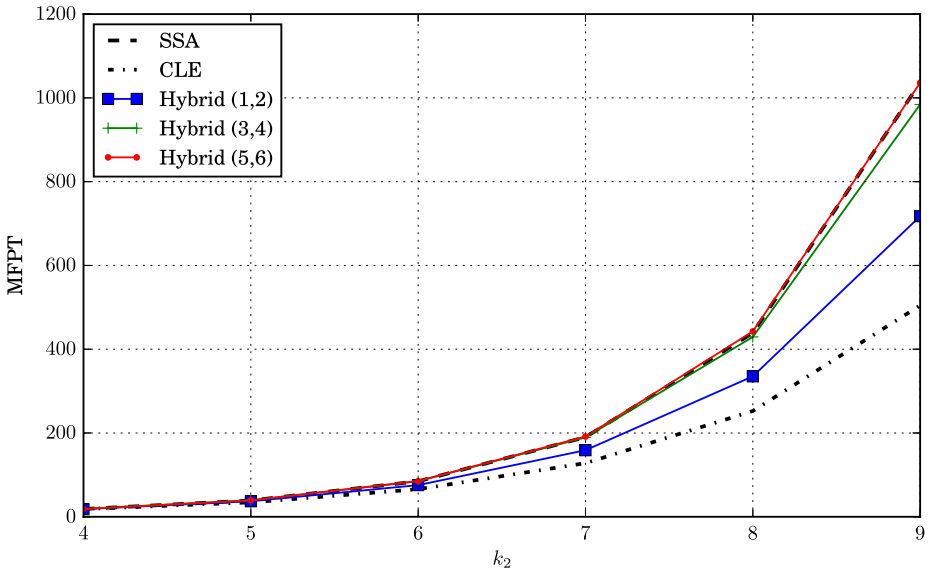
To test the hybrid scheme for MET problems, we consider the above birth–death process $\mathbf{X}(t)$ with $k_1 = 1$, so that $n_e = k_2$. We compute the mean time of the birth–death process starting from n_e to extinction. Following the discussion at the end of Section 2, the hybrid scheme is considered extinct when the process satisfies $[\![\mathbf{Z}(t)]\!] = 0$. We choose the blending functions according to (6), simulating the process for different values of I_1^i and I_2^i. Since the blending region is bounded, we can use the thinning-based Algorithm 3 to simulate the jump-diffusion process within the blending region, choosing $\Lambda_1 = k_1 I_2^1$ and $\Lambda_2 = k_2$. A timestep of $\delta t = \Delta t = 10^{-2}$ is chosen throughout.

In Fig. 6, we plot the MET of the hybrid scheme for varying k_2, each point generated from 10^5 independent realisations, and for different choices of blending regions. The MET for the CME and CLE, computed directly from (24) and (26), respectively, are shown for comparison. It is evident from the numerical experiments that the hybrid scheme provides a better approximation for the MER compared to the CLE. However, the improvement is not uniform over all timescales: the region in which the jump process is simulated must be increased to correctly capture rare events. Fig. 6 suggests that the width of the blending region also plays a role in the simulation. Indeed, a blending region with $(I_1^i, I_2^i) = (3, 5)$ appears to be sufficient to accurately estimate the MET up to $k_1 = 9$, although it is likely this approximation will break down, if k_2 is increased further. Unlike in the previous examples, we see here that the choice of blending region does play a significant role for capturing rare events, and it is evident that the choice of blending region will determine the timescale on which rare events will be observed. While this lack of a uniform approximation over all timescales is a disadvantage, the hybrid scheme provides us with an approach for improving the MET estimate obtained from the CLE, at the "cost" of having to simulate discrete jumps in (increasingly large) regions of the domain.

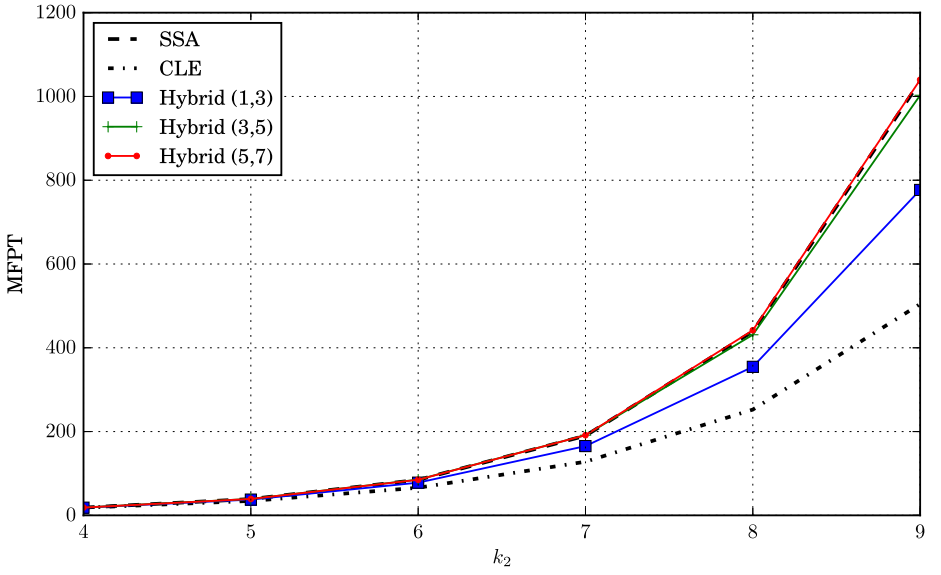
6. Conclusions

In this paper we have introduced a jump-diffusion model for simulating multiscale reaction systems efficiently while still accounting for the discrete fluctuations where necessary. Fast reactions are simulated using the CLE, while the standard discrete description is used for slow ones. Our approach involves the introduction of a set of blending functions (6) which allow one to make explicit in which regions the continuum approximation should be expected to hold.

Based on the representation of the Markov jump process as a time changed Poisson process, we described three different schemes, based on [25,1] to numerically simulate the jump-diffusion model in the three different regimes (discrete, continuous and hybrid). To demonstrate the efficacy of the schemes, we simulated equilibrium distributions of chemical systems and computed extinction times of chemical species for illustrative chemical systems. The results suggest that the proposed



(a) *Blending regions of width 1*



(b) *Blending regions of width 2*

Fig. 6. Comparison of the METs of the birth–death process (23), the corresponding CLE (25) and the hybrid scheme. The parameters used are $k_2 = 4, 5, 6, 7, 8, 9$ and $k_1 = 1$ and the process is started from $A(0) = k_2$. We use: (a) blending regions of width 1; (b) blending regions of width 2.

algorithm is robust, and is able to handle multiscale processes efficiently without the breakdown associated when using the CLE directly.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the *European Community*'s Seventh Framework Programme (*FP7/2007-2013*)/ERC *grant agreement* n° 239870. Radek Erban would also like to thank the Royal Society for a University Research Fellowship and the Leverhulme Trust for a Philip Leverhulme Prize. Andrew Duncan was also supported by the EPSRC grant EP/J009636/1. Authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme "Stochastic Dynamical Systems in Biology: Numerical Methods and Applications" where work on this paper was partially undertaken. This work was supported by EPSRC grant no. EP/K032208/1. This work was supported by a grant from the Simons Foundation.

References

[1] D.F. Anderson, A modified next reaction method for simulating chemical systems with time dependent propensities and delays, J. Chem. Phys. 127 (21) (2007) 214107.
[2] D.F. Anderson, T.G. Kurtz, Continuous time Markov chain models for chemical reaction networks, in: Design and Analysis of Biomolecular Circuits, Springer, 2011, pp. 3–42.
[3] D.F. Anderson, J.C. Mattingly, A weak trapezoidal method for a class of stochastic differential equations, Commun. Math. Sci. 9 (1) (2011).
[4] A. Angius, G. Balbo, M. Beccuti, E. Bibbona, A. Horvath, R. Sirovich, Approximate analysis of biological systems by hybrid switching jump diffusion, Theor. Comput. Sci. 587 (2015) 49–72.
[5] A. Arkin, J. Ross, H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage l-infected Escherichia coli cells, Genetics 149 (1998) 1633–1648.
[6] S. Badia, M. Parks, P. Bochev, M. Gunzburger, R. Lehoucq, On atomistic-to-continuum coupling by blending, Multiscale Model. Simul. 7 (1) (2008) 381–406.
[7] R.E. Barnhill, A.J. Worsey, Smooth interpolation over hypercubes, Comput. Aided Geom. Des. 1 (2) (1984) 101–113.
[8] J. Bezanson, S. Karpinski, V.B. Shah, A. Edelman, Julia: a fast dynamic language for technical computing, arXiv preprint, arXiv:1209.5145, 2012.
[9] Y. Cao, H. Li, L. Petzold, Efficient formulation of the stochastic simulation algorithm for chemically reacting systems, J. Chem. Phys. 121 (9) (2004) 4059–4067.
[10] S.L. Cotter, Constrained approximation of effective generators for multiscale stochastic reaction networks and application to conditioned path sampling, arXiv preprint, arXiv:1506.02446, 2015.
[11] S.L. Cotter, T. Vejchodsky, R. Erban, Adaptive finite element method assisted by stochastic simulation of chemical systems, SIAM J. Sci. Comput. 35 (1) (2013) B107–B131.
[12] A. Crudu, A. Debussche, O. Radulescu, Hybrid stochastic simplifications for multiscale gene networks, BMC Syst. Biol. 3 (1) (2009) 89.
[13] S. Dana, S. Raha, Physically consistent simulation of mesoscale chemical kinetics: the non-negative FIS-α method, J. Comput. Phys. 230 (24) (2011) 8813–8834.
[14] C.R. Doering, K.V. Sargsyan, L.M. Sander, Extinction times for birth–death processes: exact results, continuum asymptotics, and the failure of the Fokker–Planck approximation, Multiscale Model. Simul. 3 (2) (2005) 283–299.
[15] A. Duncan, S. Liao, T. Vejchodskỳ, R. Erban, R. Grima, Noise-induced multistability in chemical systems: discrete versus continuum modeling, Phys. Rev. E 91 (4) (2015) 042111.
[16] R. Erban, S.J. Chapman, P. Maini, A practical guide to stochastic simulations of reaction–diffusion processes, 35 pages, available as http://arxiv.org/abs/0704.1908, 2007.
[17] R. Erban, S.J. Chapman, I.G. Kevrekidis, T. Vejchodský, Analysis of a stochastic chemical system close to a SNIPER bifurcation of its mean-field model, SIAM J. Appl. Math. 70 (3) (2009) 984–1016.
[18] L. Ferm, P. Lötstedt, P. Sjöberg, Adaptive, conservative solution of the Fokker–Planck equation in molecular biology, Technical report 2004-054, Dept. of Information Technology, Uppsala University, Uppsala, Sweden, 2004, available at http://www.it.uu.se/research/reports/2004-054, 2004.
[19] M.B. Flegg, S.J. Chapman, R. Erban, The two-regime method for optimizing stochastic reaction–diffusion simulations, J. R. Soc. Interface 9 (70) (Oct 2011) 859–868.
[20] M.B. Flegg, S. Jonathan Chapman, L. Zheng, R. Erban, Analysis of the two-regime method on square meshes, SIAM J. Sci. Comput. 36 (3) (Jan 2014) B561–B588.
[21] B. Franz, M.B. Flegg, S.J. Chapman, R. Erban, Multiscale reaction–diffusion algorithms: PDE-assisted brownian dynamics, SIAM J. Appl. Math. 73 (3) (2013) 1224–1247.
[22] A. Ganguly, D. Altintan, H. Koeppl, Jump-diffusion approximation of stochastic reaction dynamics: error bounds and algorithms, Multiscale Model. Simul. 13 (4) (2015) 1390–1419.
[23] G.W. Gardiner, Handbook of Stochastic Processes for Physics, Chemistry and Natural Sciences, 2nd edition, Springer Verlag, 1985.
[24] M.A. Gibson, J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, J. Phys. Chem. A 104 (9) (2000) 1876–1889.
[25] D. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (25) (1977) 2340–2361.
[26] D.T. Gillespie, A rigorous derivation of the chemical master equation, Phys. A, Stat. Mech. Appl. 188 (1) (1992) 404–425.
[27] D.T. Gillespie, The chemical Langevin equation, J. Chem. Phys. 113 (1) (2000) 297–306.
[28] W.J. Gordon, Blending-function methods of bivariate and multivariate interpolation and approximation, SIAM J. Numer. Anal. 8 (1) (1971) 158–177.
[29] R. Grima, P. Thomas, A.V. Straube, How accurate are the nonlinear chemical Fokker–Planck and chemical Langevin equations?, J. Chem. Phys. 135 (8) (2011) 084103.
[30] P. Hanggi, H. Grabert, P. Talkner, H. Thomas, Bistable systems: master equation versus Fokker–Planck modeling, Phys. Rev. A 29 (1) (1984) 371.
[31] E. Haseltine, J. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics, J. Chem. Phys. 117 (2002) 6959–6969.
[32] J. Hasenauer, V. Wolf, A. Kazeroonian, F.J. Theis, Method of conditional moments (MCM) for the chemical master equation, J. Math. Biol. 69 (3) (2014) 687–735.
[33] A. Hellander, P. Lötstedt, Hybrid method for the chemical master equation, J. Comput. Phys. 227 (1) (2007) 100–122.
[34] B. Hepp, A. Gupta, M. Khammash, Adaptive hybrid simulations for multiscale stochastic reaction networks, J. Chem. Phys. 142 (3) (2015) 034118.
[35] R. Hinch, S.J. Chapman, Exponentially slow transitions on a Markov chain: the frequency of calcium sparks, Eur. J. Appl. Math. 16 (2005) 427–446.
[36] T. Jahnke, W. Huisinga, Solving the chemical master equation for monomolecular reaction systems analytically, J. Math. Biol. 54 (1) (2007) 1–26.
[37] T. Jahnke, On reduced models for the chemical master equation, Multiscale Model. Simul. 9 (4) (2011) 1646–1676.
[38] T. Jahnke, M. Kreim, Error bound for piecewise deterministic processes modeling stochastic reaction systems, Multiscale Model. Simul. 10 (4) (2012) 1119–1147.
[39] S. Kar, W. Baumann, M. Paul, J. Tyson, Exploring the roles of noise in the eukaryotic cell cycle, Proc. Natl. Acad. Sci. USA 16 (2009) 6471–6476.
[40] J. Karlsson, R. Tempone, Towards automatic global error control: computable weak error expansion for the tau-leap method, Monte Carlo Methods Appl. 17 (3) (2011) 233–278.
[41] Y.N. Kaznessis, Computational methods in synthetic biology, Biotechnol. J. 4 (10) (2009) 1392–1405.
[42] G. Klingbeil, R. Erban, M. Giles, P. Maini, STOCHSIMGPU: parallel stochastic simulation for the Systems Biology Toolbox 2 for MATLAB, Bioinformatics 27 (8) (2011) 1170–1171.
[43] G. Klingbeil, R. Erban, M. Giles, P.K. Maini, Fat versus thin threading approach on GPUs: application to stochastic simulation of chemical reactions, IEEE Trans. Parallel Distrib. Syst. 23 (2) (2012) 280–287.
[44] T.G. Kurtz, Approximation of Population Processes, vol. 36, SIAM, 1981.
[45] P.A. Lewis, G.S. Shedler, Simulation of nonhomogeneous Poisson processes by thinning, Nav. Res. Logist. Q. 26 (3) (1979) 403–413.
[46] S. Liao, T. Vejchodskỳ, R. Erban, Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks, J. R. Soc. Interface 12 (108) (2015) 20150233.

[47] D. Luc, Non-Uniform Random Variate Generation, Springer, New York, 1986.
[48] B. Melykuti, K. Burrage, K.C. Zygalakis, Fast stochastic simulation of biochemical reaction systems by alternative formulations of the chemical Langevin equation, J. Chem. Phys. 132 (16) (2010) 164109.
[49] S. Menz, J.C. Latorre, C. Schutte, W. Huisinga, Hybrid stochastic–deterministic solution of the chemical master equation, Multiscale Model. Simul. 10 (4) (2012) 1232–1262.
[50] C. Safta, K. Sargsyan, B. Debusschere, H.N. Najm, Hybrid discrete/continuum algorithms for stochastic reaction networks, J. Comput. Phys. 281 (2015) 177–198.
[51] H. Salis, Y. Kaznessis, Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions, J. Chem. Phys. 122 (5) (2005) 054103.
[52] D. Schnoerr, G. Sanguinetti, R. Grima, The complex chemical Langevin equation, J. Chem. Phys. 141 (2) (2014) 024103.
[53] P. Sjöberg, P. Lötstedt, J. Elf, Fokker–Planck approximation of the master equation in molecular biology, Comput. Vis. Sci. 12 (1) (2009) 37–50.
[54] S. Smith, C. Cianci, R. Grima, Model reduction for stochastic chemical systems with abundant species, arXiv preprint, arXiv:1510.03172, 2015.
[55] V.H. Thanh, C. Priami, Simulation of biochemical reactions with time-dependent rates by the rejection-based algorithm, J. Chem. Phys. 143 (21) (2015) 214105.
[56] N. van Kampen, Stochastic Processes in Physics and Chemistry, 3rd edition, North-Holland, Amsterdam, 2007.
[57] J. Villar, H. Kueh, N. Barkai, S. Leibler, Mechanisms of noise-resistance in genetic oscillators, Proc. Natl. Acad. Sci. USA 99 (9) (2002) 5988–5992.
[58] M. Voliotis, P. Thomas, R. Grima, C.G. Bowsher, Stochastic simulation of biomolecular networks in dynamic environments, PLoS Comput. Biol. 12 (6) (2016) e1004923.